

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:06:08 ; Search time 299.224 Seconds
(without alignments)
593.509 Million cell updates/sec

Title: US-10-010-476-10
Perfect score: 30
Sequence: 1 CTGGATCTTGGCCGCCCTTGAAATCCC 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	2	AAX40725 GC-box b-
2	30	100.0	30	2	AAX40724 GC-box a
3	30	100.0	50	2	AAX40728 GC-box c-
4	22	73.3	22	2	AAX40729 GC-box d-
5	21.6	72.0	705	6	ABT09033 Phase-1 R
6	21.6	72.0	705	12	ADG45731 Liver inf
7	21.6	72.0	4562	13	ADRI5041 Rat elect
8	20.4	68.0	424	9	ACH04308 Human cDN
9	20.4	68.0	466	8	ABZ18450 Group III
10	20	66.7	504	10	ADD34027 Mouse mit
11	20	66.7	1760	10	ABT42112 Toxicity
12	20	66.7	3374	2	AAV20872 Murine ph
13	19.6	65.3	655	6	ABT09364 Phase-1 R
14	19.6	65.3	655	12	ADG45424 Liver inf
15	19.6	65.3	4195	4	AAH98347 Human EST
16	19.6	65.3	4385	4	AAH22674 Human cDN
17	19.6	65.3	6324	3	AACT71106 Human ORF
18	19.6	65.3	50000	9	ADB16929 Human DYX
19	19.4	64.7	33	2	AAQ61741 HEV strai
20	19.4	64.7	33	2	AAT27452 HEV strai

94	18.6	62.0	1789	13	ADR24724	AdR24724 Breast ca	c 167	17.8	59.3	2061	5	AAS67616	Aas67616 DNA encod
95	18.6	62.0	21925	10	ADC87602	AdC87602 Human GPC	168	17.8	59.3	2207	3	AAC59244	Aac59244 Human sec
96	18.6	62.0	26184	10	ADC87236	AdC87236 Human GPC	169	17.8	59.3	2567	4	ABL19208	AbL19208 Drosophil
97	18.6	62.0	96599	9	AA157703	Aa157703 Human pro	c 170	17.8	59.3	3386	10	ADA53520	Ada53520 Human cod
98	18.6	62.0	96599	9	ADA02672	AdA02672 Human PRL	c 171	17.8	59.3	6260	4	AAK73034	Aak73034 Human imm
99	18.6	62.0	96599	10	ADB72410	AdB72410 Human PRL	c 172	17.8	59.3	16499	4	AAK77656	Aak77656 Human imm
100	18.6	62.0	96599	10	ADE95920	AdE95920 Human PRL	c 173	17.8	59.3	58326	12	ADO48537	AdO48537 Human can
101	18.6	62.0	110000	10	AAL52246	Aa152246 Human gen	c 174	17.8	59.3	65423	13	ADB23566	AdB23566 Human can
102	18.6	62.0	158001	12	ADL17884	Ad117884 Human pha	c 175	17.8	59.3	73995	11	ACN43986	AcN43986 Human gen
103	18.4	61.3	647	6	AQ055221	AbQ55221 Human ova	c 176	17.8	59.3	98606	11	ACN43868	AcN43868 Mouse gen
104	18.4	61.3	2066	4	ABQ52579	AbQ52579 Human ova	c 177	17.8	59.3	106315	11	ACN43966	AcN43966 Human gen
105	18.4	61.3	2391	8	ABT19233	AbT19233 Aspergill	c 178	17.8	59.3	111836	13	ABD33102	AbD33102 Murine ca
106	18.4	61.3	2391	8	ABT120455	AbT20455 Aspergill	c 179	17.8	59.3	183610	8	ACF62736	AcF62736 Murine ca
107	18.4	61.3	2391	8	ABT18639	AbT18639 Aspergill	c 180	17.8	59.3	183610	8	ADB20851	AdB20851 MRP1 base
108	18.4	61.3	2391	8	ABT21053	AbT21053 Aspergill	c 181	17.8	59.3	183610	10	ADB87940	AdB87940 Human UGT
109	18.4	61.3	4390	8	ABT18045	AbT18045 Aspergill	c 182	17.8	59.3	183610	10	ADB86923	AdB86923 Human MDR
110	18.4	61.3	4391	8	ABT19859	AbT19859 Aspergill	c 183	17.8	59.3	183610	10	ADB92114	AdB92114 Human MDR
111	18.4	61.3	5997	4	AAS32580	AaE32580 Human gen	c 184	17.8	59.3	186591	8	ACF62750	AcF62750 Cancer ba
112	18.4	61.3	6093	4	AAS32581	AaE32581 Human gen	c 185	17.8	59.3	186591	8	ADB20869	AdB20869 MRP1 base
113	18.4	61.3	8334	13	ADR84240	AdR84240 Aspergill	c 186	17.8	59.3	186591	10	ADB87958	AdB87958 Human UGT
114	18.4	61.3	94001	13	ADJ33491	AdJ33491 Human LAR	c 187	17.8	59.3	186591	10	ADB96941	AdB96941 Human MDR
115	18.4	61.3	110000	6	ABX08336	AbX08336 Continuation (7 of	c 188	17.8	59.3	186591	10	ADB92132	AdB92132 Human MDR
116	18.4	61.3	110000	12	ADJ25985	AbJ25985 Continuation (7 of	c 189	17.8	59.3	203132	13	ABD33364	AbD33364 Murine ca
117	18.4	61.3	110000	12	ADN97989	AdN97989 Continuation (7 of	c 190	17.8	59.3	208648	8	ACF62735	AcF62735 Cancer ba
118	18.4	61.3	110000	12	ADN46845	AdN46845 Continuation (5 of	c 191	17.8	59.3	208648	8	ACF62740	AcF62740 Cancer ba
119	18.4	61.3	110000	12	ADN47591	AdN47591 Continuation (5 of	c 192	17.8	59.3	208648	8	ADB20850	AdB20850 MRP1 base
120	18.4	61.3	110000	12	ADN46123	AdN46123 Continuation (5 of	c 193	17.8	59.3	208648	8	ADB20855	AdB20855 MRP1 base
121	18.4	61.3	110000	12	ADN47209	AdN47209 Continuation (17 of	c 194	17.8	59.3	208648	10	ADB87944	AdB87944 Human UGT
122	18.4	61.3	110000	12	ADN46464	AdN46464 Continuation (5 of	c 195	17.8	59.3	208648	10	ADB87939	AdB87939 Human UGT
123	18.4	61.3	110000	12	ADN47960	AdN47960 Continuation (17 of	c 196	17.8	59.3	208648	10	ADB96922	AdB96922 Human MDR
124	18.4	61.3	110000	12	ADO50281	AdO50281 Continuation (7 of	c 197	17.8	59.3	208648	10	ADB96927	AdB96927 Human MDR
125	18.4	61.3	194883	11	ACN44398	AcN44398 Human gen	c 198	17.8	59.3	208648	10	ADB92113	AdB92113 Human MDR
126	18.2	60.7	716	6	AB199929	Ab199929 Rat mucoc	c 199	17.8	59.3	208648	10	ADB92118	AdB92118 Human MDR
127	18.2	60.7	787	6	AB199930	Ab199930 Rat mucoc	c 200	17.8	59.3	214520	10	ADL13471	AdL13471 Osteoarth
128	18.2	60.7	2298	11	ADM02690	AdM02690 Human CDN	c 201	17.6	58.7	234	6	ABN19770	AbN19770 Human ORP
129	18.2	60.7	2568	6	AD27808	Ad27808 Human pro	c 202	17.6	58.7	333	4	AAI22307	AaI22307 Probe #12
130	18.2	60.7	2634	6	AD339055	Aad339055 Human 577	c 203	17.6	58.7	333	4	ABA67385	AbA67385 Human foe
131	18.2	60.7	2667	6	AD27807	Ad27807 Human pro	c 204	17.6	58.7	333	4	ABA49470	AbA49470 Human bre
132	18.2	60.7	3657	11	ACN92864	AcN92864 Breast ca	c 205	17.6	58.7	333	4	AAK41552	AaK41552 Human bon
133	18.2	60.7	5998	8	AB224582	Ab224582 Human cel	c 206	17.6	58.7	333	5	AAI07994	AaI07994 Probe #79
134	18.2	60.7	6591	12	ADQ24342	AdQ24342 Human sof	c 207	17.6	58.7	333	6	ASB15563	AsB15563 Human gen
135	18.2	60.7	15810	6	AA516995	AaE16995 Human atr	c 208	17.6	58.7	347	6	ABN20125	AbN20125 Human ORP
136	18.2	60.7	33916	13	ABD33626	AbD33626 Murine ca	c 209	17.6	58.7	396	8	ABX65117	AbX65117 Human foe
137	18.0	60.0	80	12	ADM96117	AdM96117 Rat antis	c 210	17.6	58.7	486	9	ACH44119	AcH44119 Human foe
138	18.0	60.0	347	8	AD448918	Ad448918 Human met	c 211	17.6	58.7	498	4	AAI13071	AaI13071 Probe #30
139	18.0	60.0	378	4	AAK59216	AaK59216 Human imm	c 212	17.6	58.7	498	4	ABA54770	AbA54770 Human foe
140	18.0	60.0	508	4	AAK74750	AaK74750 Human imm	c 213	17.6	58.7	498	4	ABA44316	AbA44316 Human bre
141	18.0	60.0	546	4	AB112729	Ab112729 Drosophil	c 214	17.6	58.7	498	4	AAK28495	AaK28495 Human bon
142	18.0	60.0	1177	2	AAV13960	AaV13960 Rat inter	c 215	17.6	58.7	498	5	AAI02976	AaI02976 Probe #29
143	18.0	60.0	1249	2	AAV13959	AaV13959 Rat inter	c 216	17.6	58.7	498	6	ABX03008	AbX03008 Human gen
144	18.0	60.0	1249	2	ADBS52637	AdB52637 Primary r	c 217	17.6	58.7	578	12	ACH80054	AcH80054 Human gen
145	18.0	60.0	2000	8	ADA72589	AdA72589 Rice gene	c 218	17.6	58.7	837	13	ADT44430	AdT44430 Bacterial
146	18.0	60.0	2205	12	ADN17324	AdN17324 H. annosu	c 219	17.6	58.7	1143	9	ADB06241	AdB06241 Alloiococ
147	18.0	60.0	2573	10	ABT42316	AbT42316 Toxicity	c 220	17.6	58.7	1143	9	ADB06239	AdB06239 Alloiococ
148	18.0	60.0	2606	4	AB112728	Ab112728 Drosophil	c 221	17.6	58.7	1143	9	ADB06243	AdB06243 Alloiococ
149	18.0	60.0	80077	11	ACN45040	AcN45040 Mouse gen	c 222	17.6	58.7	1679	6	ABL39772	AbL39772 Human NS
150	18.0	60.0	349980	6	ABQ081846	AbQ081846 Bifidobac	c 223	17.6	58.7	1729	13	ADT44742	AdT44742 Bacterial
151	17.8	59.3	461	5	ADL38130	AdL38130 Human ova	c 224	17.6	58.7	2331	4	ABL09127	AbL09127 Drosophil
152	17.8	59.3	461	5	ADL72995	AdL72995 Human ova	c 225	17.6	58.7	2347	4	AAH19322	AaH19322 Human ubi
153	17.8	59.3	714	9	AA162366	AaL62366 Human tet	c 226	17.6	58.7	2347	10	ABX93880	AbX93880 Human cDN
154	17.8	59.3	716	10	ADK15259	AdK15259 Urinary s	c 227	17.6	58.7	2412	4	AAF81717	AaF81717 Human pro
155	17.8	59.3	920	4	ABL19209	Ab119209 Drosophil	c 228	17.6	58.7	3660	4	AAH17633	AaH17633 Human cDN
156	17.8	59.3	1083	10	ADE57621	AdE57621 Human gen	c 229	17.6	58.7	3951	10	ADP28763	AdP28763 Human deu
157	17.8	59.3	1083	10	ADE57625	AdE57625 Human gen	c 230	17.6	58.7	4331	4	ABL09126	AbL09126 Drosophil
158	17.8	59.3	1083	10	ADE57633	AdE57633 Human gen	c 231	17.6	58.7	4483	12	ADMA4026	AdM4026 Novel hum
159	17.8	59.3	1083	10	ADE57617	AdE57617 Human gen	c 232	17.6	58.7	4483	12	ADMA4026	AdM4026 Novel hum
160	17.8	59.3	1083	10	ADE57629	AdE57629 Human gen	c 233	17.6	58.7	4659	6	ABK86140	AbK86140 cDNA enco
161	17.8	59.3	1083	10	ADE57637	AdE57637 Human gen	c 234	17.6	58.7	6408	4	AAK89966	AaK89966 Human dig
162	17.8	59.3	1149	4	AAH32523	AaH32523 Human gen	c 235	17.6	58.7	6960	6	ABL34398	AbL34398 Human imm
163	17.8	59.3	1200	10	ADK15260	AdK15260 Urinary s	c 236	17.6	58.7	9929	4	AAK81767	AaK81767 Human imm
164	17.8	59.3	1340	10	ADE28242	AdE28242 Human MDD	c 237	17.6	58.7	10241	4	AAK33394	AaK33394 DNA encod
165	17.8	59.3	1470	10	ADE72870	AdE72870 Human end	c 238	17.6	58.7	10241	4	AAK83768	AaK83768 Human imm
166	17.8	59.3	1638	10	ADE72872	AdE72872 Human end	c 239	17.6	58.7	16682	4	AAK79798	AaK79798 Human imm

C 240	17.6	58.7	16682	4	AAK70010	Human imm	Aak70010 Human imm
C 241	17.6	58.7	21436	4	AAK79799	Human imm	Aak79799 Human imm
C 242	17.6	58.7	21436	4	AAK70011	Human imm	Aak70011 Human imm
C 243	17.6	58.7	23670	6	ABA97077	Tomato He	Aba97077 Tomato He
C 244	17.6	58.7	88939	13	ABD32535_3	Continuation (4 of	Abd32535_3 Continuation (4 of
C 245	17.6	58.7	89182	11	ACN44582_3	Continuation (4 of	Acn44582_3 Continuation (4 of
C 246	17.6	58.7	90104	4	ABL12402	Drosophill	Abl12402 Drosophill
C 247	17.6	58.7	110000	4	AAI99682_09	Continuation (10 o	Aai99682_09 Continuation (10 o
C 248	17.6	58.7	110000	4	AAI99683_09	Continuation (10 o	Aai99683_09 Continuation (10 o
C 249	17.6	58.7	110000	9	ADB12064_00	Allolococ	Adb12064 Allolococ
C 250	17.6	58.7	139573	10	ADH58564	Human Na+	Adh58564 Human Na+
C 251	17.6	58.7	156416	13	ABD32817	Human can	Abd32817 Human can
C 252	17.4	58.0	32	2	AAT80116	Primer CP	Aat80116 Primer CP
C 253	17.4	58.0	32	2	AAT83162	HCV prote	Aat83162 HCV prote
C 254	17.4	58.0	327	4	AAAS35821	Human car	Aas35821 Human car
C 255	17.4	58.0	327	4	AAAS36742	Human car	Aas36742 Human car
C 256	17.4	58.0	327	10	ADBA6515	Human car	Ada6515 Human car
C 257	17.4	58.0	327	10	ADBA7436	Human car	Ada7436 Human car
C 258	17.4	58.0	327	13	ADJ08854	Human car	Adj08854 Human car
C 259	17.4	58.0	327	13	ADJ07933	Human car	Adj07933 Human car
C 260	17.4	58.0	500	9	ACH41729	Human foe	Ach41729 Human foe
C 261	17.4	58.0	571	9	ACH41019	Human foe	Ach41019 Human foe
C 262	17.4	58.0	715	2	ADRO2017	A. gossyp	Adr02017 A. gossyp
C 263	17.4	58.0	717	10	ADB69642	C. neofor	Adb69642 C. neofor
C 264	17.4	58.0	783	11	ABD17569	Pseudomon	Abd17569 Pseudomon
C 265	17.4	58.0	784	6	ABQ48044	Oligonuel	Abq48044 Oligonuel
C 266	17.4	58.0	784	6	ABQ48045	Oligonuel	Abq48045 Oligonuel
C 267	17.4	58.0	823	12	ACH89781	Human gen	Ach89781 Human gen
C 268	17.4	58.0	869	6	ABQ48734	Oligonuel	Abq48734 Oligonuel
C 269	17.4	58.0	869	6	ABQ48735	Oligonuel	Abq48735 Oligonuel
C 270	17.4	58.0	873	10	ABQ807468	Novel cod	Abq807468 Novel cod
C 271	17.4	58.0	915	5	AAAS70780	DNA encod	Aas70780 DNA encod
C 272	17.4	58.0	932	9	ADAI14449	Mouse spe	Ada14449 Mouse spe
C 273	17.4	58.0	1002	10	ADB69281	C. neofor	Adb69281 C. neofor
C 274	17.4	58.0	1024	3	AAZ97168	Human pro	Aaz97168 Human pro
C 275	17.4	58.0	1056	5	AAAS70235	DNA encod	Aas70235 DNA encod
C 276	17.4	58.0	1405	8	ABAS59882	Oestrogen	Abas59882 Oestrogen
C 277	17.4	58.0	1681	13	ADSA48524	Bacterial	Adsa48524 Bacterial
C 278	17.4	58.0	1888	4	ABLO2055	Drosophill	Ablo2055 Drosophill
C 279	17.4	58.0	2091	2	AAK90698	Human neu	Aak90698 Human neu
C 280	17.4	58.0	2091	6	ABK90730	cDNA enco	Abk90730 cDNA enco
C 281	17.4	58.0	2502	2	AAK90698	Human neu	Aak90698 Human neu
C 282	17.4	58.0	2502	6	ABK90731	cDNA enco	Abk90731 cDNA enco
C 283	17.4	58.0	2534	4	ABLI12541	Drosophill	Abli12541 Drosophill
C 284	17.4	58.0	3002	10	ADB68920	C. neofor	Adb68920 C. neofor
C 285	17.4	58.0	3833	11	ACN44319	Human mRN	Acn44319 Human mRN
C 286	17.4	58.0	3975	13	ADR14674	Human NF-	Adr14674 Human NF-
C 287	17.4	58.0	4153	6	ABK34686	Human cDN	Abk34686 Human cDN
C 288	17.4	58.0	4652	4	ABLO2054	Drosophill	Ablo2054 Drosophill
C 289	17.4	58.0	4834	13	ADS48705	Bacterial	Ads48705 Bacterial
C 290	17.4	58.0	4923	13	ADR06956	Full leng	Adr06956 Full leng
C 291	17.4	58.0	5917	2	AAV17090	Porcine a	Aav17090 Porcine a
C 292	17.4	58.0	5917	10	ABS56248	Human nov	Abs56248 Human nov
C 293	17.4	58.0	6941	10	ADC30320	Human PRO	Adc30320 Human PRO
C 294	17.4	58.0	6996	12	ADOL19684	Human PRO	Adol19684 Human PRO
C 295	17.4	58.0	6996	13	ADP54810	PRO polyp	Adp54810 PRO polyp
C 296	17.4	58.0	6996	13	ADP24218	PRO polyp	Adp24218 PRO polyp
C 297	17.4	58.0	9320	4	ABLI12540	Drosophill	Abli12540 Drosophill
C 298	17.4	58.0	20000	12	ADN58769	Human B7H	Adn58769 Human B7H
C 299	17.4	58.0	23833	11	ACN44318	Human gen	Acn44318 Human gen
C 300	17.4	58.0	31527	9	ADA02945	Mouse Nek	Ada02945 Mouse Nek
C 301	17.4	58.0	31527	10	ADB72683	Mouse Nek	Adb72683 Mouse Nek
C 302	17.4	58.0	31527	10	ADC85425	Mouse Nek	Adc85425 Mouse Nek
C 303	17.4	58.0	31527	12	ADM74540	Murine ca	Adm74540 Murine ca
C 304	17.4	58.0	92861	12	ADQ97843	Mouse can	Adq97843 Mouse can
C 305	17.4	58.0	110000	11	ACN43984_2	Continuation (3 of	Acn43984_2 Continuation (3 of
C 306	17.4	58.0	256493	11	ACN44514	Human gen	Acn44514 Human gen
C 307	17.2	57.3	25	4	AAAS00105	Apple deg	Aas00105 Apple deg
C 308	17.2	57.3	216	6	ABN19498	Human ORF	Abn19498 Human ORF
C 309	17.2	57.3	324	6	ABN79462	Human ORF	Abn79462 Human ORF
C 310	17.2	57.3	388	4	AAAS30821	Human cDN	Aas30821 Human cDN
C 311	17.2	57.3	411	5	AAAS65553	Novel hum	Aas65553 Novel hum
C 312	17.2	57.3	422	4	AAK43752	Human bon	Aak43752 Human bon
C 313	17.2	57.3	422	4	AAK30749	Human bon	Aak30749 Human bon
C 314	17.2	57.3	455	9	ACH47251	Human inf	Ach47251 Human inf
C 315	17.2	57.3	482	5	ABV13574	Human pro	Abv13574 Human pro
C 316	17.2	57.3	550	4	AAI21112	Human bre	Aai21112 Human bre
C 317	17.2	57.3	587	5	AAAS34163	Human cDN	Aas34163 Human cDN
C 318	17.2	57.3	675	10	ADD33429	Mouse mit	Add33429 Mouse mit
C 319	17.2	57.3	762	13	ADS46365	Bacterial	Ads46365 Bacterial
C 320	17.2	57.3	818	13	ADR65453	Cotton CD	Adr65453 Cotton CD
C 321	17.2	57.3	1050	10	ACC61690	Gene sequ	Acc61690 Gene sequ
C 322	17.2	57.3	1050	10	ADK64121	Disease t	Adk64121 Disease t
C 323	17.2	57.3	1050	13	ADT47205	Bacterial	Adt47205 Bacterial
C 324	17.2	57.3	1126	5	AAAS67517	DNA encod	Aas67517 DNA encod
C 325	17.2	57.3	1128	10	ABZ76704	WSSV 664	Abz76704 WSSV 664
C 326	17.2	57.3	1173	5	AAAS65171	DNA encod	Aas65171 DNA encod
C 327	17.2	57.3	1173	5	AAAS88302	DNA encod	Aas88302 DNA encod
C 328	17.2	57.3	1287	13	ADS57937	Bacterial	Ads57937 Bacterial
C 329	17.2	57.3	1319	6	ABS55003	Human NKX	Abs55003 Human NKX
C 330	17.2	57.3	1567	8	ACA35376	Prokaryot	Aca35376 Prokaryot
C 331	17.2	57.3	1641	10	ACC61606	Gene sequ	Acc61606 Gene sequ
C 332	17.2	57.3	1641	10	ADK64369	Disease t	Adk64369 Disease t
C 333	17.2	57.3	1994	12	ADQ64641	Novel hum	Adq64641 Novel hum
C 334	17.2	57.3	2228	4	ABL27982	Drosophill	Abi27982 Drosophill
C 335	17.2	57.3	2279	13	ACN43393	Human dia	Acn43393 Human dia
C 336	17.2	57.3	2337	13	ACN43395	Human dia	Acn43395 Human dia
C 337	17.2	57.3	2470	13	ACN43394	Human dia	Acn43394 Human dia
C 338	17.2	57.3	2555	13	ACN43392	Human dia	Acn43392 Human dia
C 339	17.2	57.3	2612	12	ADQ64123	Novel hum	Adq64123 Novel hum
C 340	17.2	57.3	2731	10	ADB62130	Human cDN	Adb62130 Human cDN
C 341	17.2	57.3	3106	4	AAK94703	Human ful	Aak94703 Human ful
C 342	17.2	57.3	3106	12	ADL31699	Full leng	Adl31699 Full leng
C 343	17.2	57.3	3215	4	ABL22840	Drosophill	Abi22840 Drosophill
C 344	17.2	57.3	3237	5	AAAS87056	DNA encod	Aas87056 DNA encod
C 345	17.2	57.3	3620	10	ADA53334	Human cod	Ada53334 Human cod
C 346	17.2	57.3	3837	13	ADS96465	Drosophill	Ads96465 Drosophill
C 347	17.2	57.3	4030	4	ABL30131	Drosophill	Abi30131 Drosophill
C 348	17.2	57.3	4346	13	ADR07203	Full leng	Adr07203 Full leng
C 349	17.2	57.3	4545	10	ADES2577	Human SQV	Ade2577 Human SQV
C 350	17.2	57.3	4545	13	ABD33445	Human can	Abd33445 Human can
C 351	17.2	57.3	4563	3	AAAS37093	Human PRO	Aas37093 Human PRO
C 352	17.2	57.3	4563	4	AAAS4387	Probe #38	Aaf54387 Probe #38
C 353	17.2	57.3	4563	4	AAAS46080	Human DNA	Aas46080 Human DNA
C 354	17.2	57.3	4563	8	ACA89530	cDNA enco	ACA89530 cDNA enco
C 355	17.2	57.3	4563	8	ACA73540	Human sec	Aca73540 Human sec
C 356	17.2	57.3	4563	8	ACA05855	Human sec	Aca05855 Human sec
C 357	17.2	57.3	4563	8	ACA66689	cDNA enco	Aca66689 cDNA enco
C 358	17.2	57.3	4563	8	ACF20264	Human sec	Acf20264 Human sec
C 359	17.2	57.3	4563	8	ACF19650	Human sec	Acf19650 Human sec
C 360	17.2	57.3	4563	8	ACD21938	Human sec	Acd21938 Human sec
C 361	17.2	57.3	4563	8	ACF13103	Human sec	Acf13103 Human sec
C 362	17.2	57.3	4563	8	ACD25206	Human sec	Acd25206 Human sec
C 363	17.2	57.3	4563	8	ACF00255	Human sec	Acf00255 Human sec
C 364	17.2	57.3	4563	8	ACA72312	Novel hum	Aca72312 Novel hum
C 365	17.2	57.3	4563	8	ACD04836	Novel hum	Acd04836 Novel hum
C 366	17.2	57.3	4563	8	ACD18297	Human sec	Acd18297 Human sec
C 367	17.2	57.3	4563	8	ACD08304	Human sec	Acd08304 Human sec
C 368	17.2	57.3	4563	8	ACA88738	Novel hum	Aca88738 Novel hum
C 369	17.2	57.3	4563	8	ACA70180	Human sec	Aca70180 Human sec
C 370	17.2	57.3	4563	8	ACD12402	Novel hum	Acd12402 Novel hum
C 371	17.2	57.3	4563	8	ACC74317	Human sec	Acc74317 Human sec
C 372	17.2	57.3	4563	8	ACD15945	Human sec	Acd15945 Human sec
C 373	17.2	57.3	4563	8	ACD25513	Novel hum	Acd25513 Novel hum
C 374	17.2	57.3	4563	8	ACD17990	Human sec	Acd17990 Human sec
C 375	17.2	57.3	4563	8	ACC88277	Human sec	Acc88277 Human sec
C 376	17.2	57.3	4563	8	ACD21631	Human sec	Acd21631 Human sec
C 377	17.2	57.3	4563	8	ACD18698	Human sec	Acd18698 Human sec
C 378	17.2	57.3	4563	8	ABX98308	Human cDN	Abx98308 Human cDN
C 379	17.2	57.3	4563	8	ACD14059	Human PRO	Acd14059 Human PRO
C 380	17.2	57.3	4563	8	ACD09839	Human sec	Acd09839 Human sec
C 381	17.2	57.3	4563	8	ACC88584	Human sec	Acc88584 Human sec
C 382	17.2	57.3	4563	8	ACD21324	Human sec	Acd21324 Human sec
C 383	17.2	57.3	4563	8	ABX75696	Human cDN	Abx75696 Human cDN
C 384	17.2	57.3	4563	8	ABX97899	Human PRO	Abx97899 Human PRO
C 385	17.2	57.3	4563	8	ACA97375	Novel hum	Aca97375 Novel hum

386	17.2	57.3	4563	8	ACA57838	Human PRO	459	17.2	57.3	4563	9	ACF10176	Human sec
387	17.2	57.3	4563	8	ACD14366	Human PRO	460	17.2	57.3	4563	9	ACF78069	Human sec
388	17.2	57.3	4563	8	ACC91149	Human sec	461	17.2	57.3	4563	9	ACD46774	Human sec
389	17.2	57.3	4563	8	ACC88891	Human sec	462	17.2	57.3	4563	9	ACD49537	Human sec
390	17.2	57.3	4563	8	ACD07088	Human PRO	463	17.2	57.3	4563	9	ACF28304	Human sec
391	17.2	57.3	4563	8	ACA67539	Human PRO	464	17.2	57.3	4563	9	ACD88994	Human PRO
392	17.2	57.3	4563	8	ACC81594	Human sec	465	17.2	57.3	4563	9	ACD84389	Human PRO
393	17.2	57.3	4563	8	ACC89198	Human sec	466	17.2	57.3	4563	9	ACD99163	cDNA enco
394	17.2	57.3	4563	8	ACC86554	Human sec	467	17.2	57.3	4563	9	ADA78063	Human sec
395	17.2	57.3	4563	8	ACC89812	Human sec	468	17.2	57.3	4563	9	ACF48905	Human sec
396	17.2	57.3	4563	8	ACC92991	Human sec	469	17.2	57.3	4563	9	ACD09225	Human sec
397	17.2	57.3	4563	8	ACA72619	Human PRO	470	17.2	57.3	4563	9	ACF12018	Human sec
398	17.2	57.3	4563	8	ACA89137	Human sec	471	17.2	57.3	4563	9	ACF41252	Human sec
399	17.2	57.3	4563	8	ACA69873	Human sec	472	17.2	57.3	4563	9	ACF15866	Human sec
400	17.2	57.3	4563	8	ACA97016	Novel hum	473	17.2	57.3	4563	9	ACF16173	Human sec
401	17.2	57.3	4563	8	ACA91012	Novel hum	474	17.2	57.3	4563	9	ACD32000	Human sec
402	17.2	57.3	4563	8	ACA70794	Human sec	475	17.2	57.3	4563	9	ACF18808	Human sec
403	17.2	57.3	4563	8	ACA95304	Novel hum	476	17.2	57.3	4563	9	ACF09255	Human sec
404	17.2	57.3	4563	8	ACC86247	Human sec	477	17.2	57.3	4563	9	ACF78376	Human sec
405	17.2	57.3	4563	8	ACC90119	Human sec	478	17.2	57.3	4563	9	ACF51975	Human sec
406	17.2	57.3	4563	8	ACD12727	Human sec	479	17.2	57.3	4563	9	ACF26462	Human sec
407	17.2	57.3	4563	8	ACF19957	Human sec	480	17.2	57.3	4563	9	ACF24255	Human sec
408	17.2	57.3	4563	8	ABX76901	Human PRO	481	17.2	57.3	4563	9	ACF63566	Human sec
409	17.2	57.3	4563	8	ACA73233	Novel hum	482	17.2	57.3	4563	9	ACF50440	Human sec
410	17.2	57.3	4563	8	ACA68776	Novel hum	483	17.2	57.3	4563	9	ACH07911	Human sec
411	17.2	57.3	4563	8	ACA74620	cDNA enco	484	17.2	57.3	4563	9	ACF13717	Human sec
412	17.2	57.3	4563	8	ACA70487	Human sec	485	17.2	57.3	4563	9	ACD41643	Human sec
413	17.2	57.3	4563	8	ACD14673	Human PRO	486	17.2	57.3	4563	9	ACF32056	Human sec
414	17.2	57.3	4563	8	ACA68345	Novel hum	487	17.2	57.3	4563	9	ACF23334	Human sec
415	17.2	57.3	4563	8	ABX98810	Novel hum	488	17.2	57.3	4563	9	ACF40024	Human sec
416	17.2	57.3	4563	8	ACC81287	Human sec	489	17.2	57.3	4563	9	ACD45546	Human sec
417	17.2	57.3	4563	8	ACA95611	Novel hum	490	17.2	57.3	4563	9	ACF53203	Human sec
418	17.2	57.3	4563	8	ACD04529	Novel hum	491	17.2	57.3	4563	9	ACF27383	Human sec
419	17.2	57.3	4563	8	ACC87970	Human sec	492	17.2	57.3	4563	9	ACF45221	Human sec
420	17.2	57.3	4563	8	ACF12632	Human sec	493	17.2	57.3	4563	9	ACF29839	Human sec
421	17.2	57.3	4563	8	ACA96347	Human PRO	494	17.2	57.3	4563	9	ACD89915	Human sec
422	17.2	57.3	4563	8	ACA65121	Human PRO	495	17.2	57.3	4563	9	ACD84696	Human PRO
423	17.2	57.3	4563	8	ACA73847	Human sec	496	17.2	57.3	4563	9	ACD98856	cDNA enco
424	17.2	57.3	4563	8	ACA74259	Novel hum	497	17.2	57.3	4563	9	ACF77148	Human sec
425	17.2	57.3	4563	8	ACA96654	Human PRO	498	17.2	57.3	4563	9	ACF76841	Human sec
426	17.2	57.3	4563	8	ACD10760	cDNA enco	499	17.2	57.3	4563	9	ACF49826	Human sec
427	17.2	57.3	4563	8	ACC91456	Human sec	500	17.2	57.3	4563	9	ACF50133	Human sec
428	17.2	57.3	4563	8	ACD02791	cDNA enco							
429	17.2	57.3	4563	8	ACC87356	Human sec							
430	17.2	57.3	4563	8	ACC85940	Human sec							
431	17.2	57.3	4563	8	ACA65428	Human PRO							
432	17.2	57.3	4563	8	ACA94245	Human sec							
433	17.2	57.3	4563	8	ACA97989	Human PRO							
434	17.2	57.3	4563	8	ACA91491	Novel hum							
435	17.2	57.3	4563	8	ACA90705	Novel hum							
436	17.2	57.3	4563	8	ACD16252	Human sec							
437	17.2	57.3	4563	8	ACD17413	Human sec							
438	17.2	57.3	4563	8	ACC92070	Human sec							
439	17.2	57.3	4563	8	ACA74927	cDNA enco							
440	17.2	57.3	4563	8	ACA91798	Human PRO							
441	17.2	57.3	4563	8	ACA71442	Human sec							
442	17.2	57.3	4563	8	ACC90842	Human sec							
443	17.2	57.3	4563	8	ACA65852	cDNA enco							
444	17.2	57.3	4563	8	ACA94997	cDNA enco							
445	17.2	57.3	4563	8	ACD16559	Human sec							
446	17.2	57.3	4563	8	ACD15638	Human sec							
447	17.2	57.3	4563	8	ABX16741	Human cDN							
448	17.2	57.3	4563	9	ACA97682	Human PRO							
449	17.2	57.3	4563	9	ACA99131	Novel hum							
450	17.2	57.3	4563	9	ACC91763	Human sec							
451	17.2	57.3	4563	9	ACD11174	Novel hum							
452	17.2	57.3	4563	9	ACD15024	Human sec							
453	17.2	57.3	4563	9	ACD11788	Human sec							
454	17.2	57.3	4563	9	ACC95917	Human sec							
455	17.2	57.3	4563	9	ACF16480	Human sec							
456	17.2	57.3	4563	9	ACF02598	Human sec							
457	17.2	57.3	4563	9	ACF02905	Human sec							
458	17.2	57.3	4563	9	ACF21492	Human sec							

ALIGNMENTS

RESULT 1

AAK40725
ID AAK40725 standard; DNA; 30 BP.

AC AAK40725;

XX 16-JUN-1999 (first entry)

XX GC-box b-MET nucleotide sequence.

KW DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;

KW allosteric site; synthetic inhibitor; cancer cell; proliferation;

KW development disorder; Huntington's disease; Down's syndrome; Hox gene;

XX OS

XX Synthetic.

XX WO9912027-A1.

XX 11-MAR-1999.

XX 12-JUN-1998; 98WO-US012351V

XX 29-AUG-1997; 97US-0057411P.

XX (REGC) UNIV CALIFORNIA.

XX PI Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Claim 10; Page 39; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 30 BP; 3 A; 14 C; 5 G; 8 T; 0 U; 0 Other;
 SQ Query Match 100.0%; Score 30; DB 2; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.0092;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
 DB 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
 RESULT 2
 AAX40724/c
 ID AAX40724 standard; DNA; 30 BP.
 XX AC AAX40724;
 XX 16-JUN-1999 (first entry)
 XX GC-box a nucleotide sequence of the specification.
 XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 XX Synthetic.
 XX WO9912027-A1.
 XX 11-MAR-1999.
 XX 12-JUN-1998; 98WO-US012351.
 XX 29-AUG-1997; 97US-0057411P.
 XX (RECC) UNIV CALIFORNIA.
 XX Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Example 2; Page 39; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit

CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 30 BP; 8 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
 SQ Query Match 100.0%; Score 30; DB 2; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.0092;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
 DB 30 CTGGATCCTTGGCCCGCCCTTGAATTCCC 1
 RESULT 3
 AAX40728
 ID AAX40728 standard; DNA; 50 BP.
 XX AC AAX40728;
 XX 16-JUN-1999 (first entry)
 XX GC-box c-MET nucleotide sequence.
 XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 XX Synthetic.
 XX WO9912027-A1.
 XX 11-MAR-1999.
 XX 12-JUN-1998; 98WO-US012351.
 XX 29-AUG-1997; 97US-0057411P.
 XX (RECC) UNIV CALIFORNIA.
 XX Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Claim 10; Page 63; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 50 BP; 8 A; 27 C; 5 G; 10 T; 0 U; 0 Other;
 SQ Query Match 100.0%; Score 30; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 0.0096;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
 DB 11 CTGGATCCTTGGCCCGCCCTTGAATTCCC 40

```

RESULT 4
AAX40729
ID AAX40729 standard; DNA; 22 BP.
XX
XX AAX40729;
AC
XX
XX 16-JUN-1999 (first entry)
XX
XX GC-box d-MET nucleotide sequence.
XX
XX DNA cytosine methyltransferase; DCM7ase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
XX ss.
XX Synthetic.
OS
XX
XX WO9912027-A1.
XX
XX 11-MAR-1999.
PD
XX
XX 12-JUN-1998; 98WO-US012351.
PF
XX
XX 29-AUG-1997; 97US-0057411P.
PR
XX (REGC ) UNIV CALIFORNIA.
PA
XX Reich NO, Flynn J;
PI
XX
XX WPI; 1999-205256/17.
DR
XX
XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
PT used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
XX Claim 10; Page 63; 114pp; English.
XX
XX The specification describes a synthetic oligonucleotide comprising a C-5
CC methylcytosine which recognizes and binds an allosteric site on DNA
CC cytosine methyltransferase (DCM7ase) ; thereby modulating DCM7ase
CC activity associated with the allosteric site. The synthetic inhibitor can
CC be used to inhibit methylation of DNA. It can also be used to inhibit
CC proliferation of cancer cells. The inhibitor can also be used to treat a
CC disorder of development, which is linked to a genetic locus regulated by
CC methylation, such as Huntington's disease, Down's syndrome, and disorders
CC associated with a Hox gene. The present sequence appears in the
CC specification
XX
XX Sequence 22 BP; 3 A; 10 C; 3 G; 6 T; 0 U; 0 Other;
SQ
Query Match 73.3%; Score 22; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ATCCTTGCCCGCCCTTGAAT 26
Db 1 ATCCTTGCCCGCCCTTGAAT 22

RESULT 5
ABT09033
ID ABT09033 standard; DNA; 705 BP.
XX
XX AC ABT09033;
XX
XX 05-DEC-2002 (first entry)
XX
XX Phase-1 Rat CT gene SEQ ID No 121.
DE
XX Rat; toxicity study; rat toxic response gene; toxicological response;
KW drug development; phase-1 rat CT gene; ds.
XX

```

```

OS Rattus sp.
XX
XX WO200266682-A2.
XX
XX 29-AUG-2002.
PD
XX
XX 29-JAN-2002; 2002WO-US002935.
XX
XX 29-JAN-2001; 2001US-0264933P.
PR
XX 26-JUL-2001; 2001US-0306161P.
XX
XX (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
PA
XX Farris G, Hicken SH, Farr SB;
XX
XX WPI; 2002-674961/72.
DR
XX
XX Evaluating the toxicity of an agent, useful in drug development or in
PT determining toxicological responses to a new drug, by determining the
PT expression of rat toxicologically relevant genes in the test animal in
PT response to the test agent.
XX
XX Disclosure; Page 135; 388pp; English.
XX
XX The invention relates to a method used for evaluating the toxicity of an
CC agent comprising determining the expression of a rat toxic response
CC gene(s) in the test animal in response to the agent. The method is useful
CC in drug development, particularly for conducting toxicity studies and
CC analysis before a new drug or compound is approved for human consumption
CC or use. The method is also useful in determining toxicological responses
CC to a new drug. This polynucleotide sequence represents a phase-1 rat CT
CC gene of the invention
XX
XX Sequence 705 BP; 180 A; 187 C; 181 G; 155 T; 0 U; 2 Other;
SQ
Query Match 72.0%; Score 21.6; DB 6; Length 705;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CTGATCCTTGCCCGCCCTTGAATTC 28
Db 185 CTGATCCTTGCCCGCCCTTGAATTC 212

RESULT 6
ADG45731
ID ADG45731 standard; DNA; 705 BP.
XX
XX AC ADG45731;
XX
XX 26-FEB-2004 (first entry)
XX
XX Liver inflammatory predictive gene related DNA sequence.
DE
XX liver; liver toxicity; liver toxicity predictive gene;
KW liver inflammation predictive gene; inflammation; gene; ds; rat.
XX
XX Rattus norvegicus.
OS
XX
XX WO2003095624-A2.
XX
XX 20-NOV-2003.
PD
XX
XX 09-MAY-2003; 2003WO-US014832.
PF
XX
XX 10-MAY-2002; 2002US-0379831P.
PR
XX
XX (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
PA
XX Kier L, Nolan TD, Sankar U, Derbel M;
PI
XX WPI; 2004-042456/04.
DR
XX

```

PT Predicting the liver toxicity in an individual to an agent by measuring
PT the expression of one or more liver toxicity predictive genes in the
XX sample from the individual treated with the agent.
XX
XX Claim 7; Fig 4; 336pp; English.
XX
CC The present invention describes a method for predicting the liver
CC toxicity in an individual to an agent. The method comprises: (a)
CC obtaining a biological sample from the individual treated with the agent;
CC (b) measuring the expression of one or more liver toxicity predictive
CC genes in the sample; and (c) using the test expression profile with a set
CC of reference expression profiles in a Predictive Model to determine
CC whether the agent will induce liver toxicity in the individual. Also
CC described: (1) predicting the liver toxicity of an agent using an in
CC vitro system; (2) predicting the liver toxicity in a biological sample
CC from an individual, in vitro cell cultures or explants to an agent via a
CC programmable machine; (3) a computer program product for enabling a
CC computer to perform Predictive Model analysis for liver toxicity on a
CC biological sample from an individual, in vitro cell cultures or explants
CC to an agent; (4) a computer system adopted to predict liver toxicity in a
CC biological system from an individual, in vitro cell cultures or explants
CC to an agent; (5) a computer program product for predicting liver toxicity
CC from a test sample expression profile; (6) mining genes predictive of
CC liver toxicity; (7) an integrated system for predicting liver toxicity;
CC and (8) identifying one or more liver inflammation predictive genes. The
CC method is useful for predicting the liver toxicity in an individual to an
CC agent. The present sequence represents a liver toxicity predictive gene
CC sequence that represents a 24 hour combo 3 gene, which is specifically
CC claimed in the exemplification of the present invention.
XX
XX Sequence 705 BP; 180 A; 187 C; 181 G; 155 T; 0 U; 2 Other;
XX
XX Query Match 72.0%; Score 21.6; DB 12; Length 705;
XX Best Local Similarity 85.7%; Pred. No. 35;
XX Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 CTGGATCCTTGGCCCGCCCTTGGAATTC 28
XX |||||
XX DB 185 CTGGATCCTTGGCCCGCCACCCCTGAGTTC 212
XX
XX RESULT 7
XX ADR15041
XX ID ADR15041 standard; DNA; 4562 BP.
XX AC ADR15041;
XX
XX DT 21-OCT-2004 (first entry)
XX
XX DE Rat electroconvulsive seizure (EC) signature gene seqid 74.
XX
XX KW neuroleptic; nootropic; antidepressant; eating disorder; tranquilizer;
XX neuropsychiatric disorder; signature gene; electroconvulsive seizure;
XX ECS; schizophrenia; autism; major depressive disorder; MDD;
XX bipolar affective disorder; BAD; psychotic depression; anxiety disorder;
XX eating disorder; attention deficit hyperactivity disorder;
XX neuropsychiatric drug; gene; ss; rat.
XX
XX OS Rattus norvegicus.
XX
XX PN US2004152107-A1.
XX
XX PD 05-AUG-2004.
XX
XX PF 18-SEP-2003; 2003US-00664705.
XX
XX PR 18-SEP-2002; 2002US-0411718P.
XX PR 09-DEC-2002; 2002US-0431882P.
XX PR 18-JUN-2003; 2003US-0479970P.
XX
XX (ALTA/) ALTAR C A.
XX (LAEN/) LAENG P.
XX (YOUN/) YOUNG T A.

PA (CHAR/) CHARLES V D.
XX Altar CA, Laeng P, Young TA, Charles VD;
XX WPI; 2004-580183/56.
XX
XX Identifying compound to treat neuropsychiatric disorder, by contacting
XX cell with test compound, determining expression of signature genes by
XX cell, comparing expression of signature genes to expression in cell not
XX contacted with compound.
XX
XX Claim 1; SEQ ID NO 74; 186pp; English.
XX
XX The invention describes a method of identifying (M1) a compound to treat
XX a neuropsychiatric disorder. The method involves contacting a cell with a
XX test compound, determining the expression of one or more signature genes
XX by the cell, each signature gene comprising a nucleic acid that
XX hybridizes to a nucleic acid chosen from 154 fully defined sequences (S1)
XX of e.g., 4154, 1582, 645, 367, 405, 1080, 1159, 539, 465, 2002, 2469,
XX 3165, 421, 2957, 644, 2562, 2858, 1194 and 1227 base pairs as given
XX in the specification and their complements, and comparing determined
XX expression of one or more signature genes to expression in cell not
XX contacted with test compound. Also described are: selecting (M2) one or
XX more signature genes that are indicative of an effective therapy for
XX treating a neuropsychiatric disorder; and a kit (I) for detecting an
XX electroconvulsive seizure (ECS) gene signature, comprising several
XX oligonucleotides, each of which is capable of specifically hybridizing to
XX a different ECS signature gene. Therapeutic methods, which use ECS gene
XX signatures and ECS signature genes to treat, ameliorate or prevent
XX neuropsychiatric disorders, is also disclosed. (M1) is useful for
XX identifying a compound to treat a neuropsychiatric disorder chosen from
XX schizophrenia, autism, major depressive disorder (MDD), bipolar affective
XX disorder (BAD) and psychotic depression. (M1) is useful for identifying
XX compounds for treating neuropsychiatric disorders such as anxiety
XX disorders, eating disorders and attention deficit hyperactivity disorder.
XX The ECS signature genes of (M1) are useful in diagnostic and prognostic
XX methods of identifying people who have or who are susceptible to
XX neuropsychiatric disorders such as MDD and BAD. (M1) enables multi-
XX parameter high throughput screening of compounds for treating
XX neuropsychiatric disorders; screening of large number of candidate
XX compounds for treating neuropsychiatric disorder and enables cost-
XX effective and rapid screening of compounds useful as neuropsychiatric
XX drugs. This sequence represents an electroconvulsive seizure (ECS)
XX signature gene.
XX
XX Sequence 4562 BP; 981 A; 1273 C; 1195 G; 1113 T; 0 U; 0 Other;
XX
XX Query Match 72.0%; Score 21.6; DB 13; Length 4562;
XX Best Local Similarity 85.7%; Pred. No. 41;
XX Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 CTGGATCCTTGGCCCGCCCTTGGAATTC 28
XX |||||
XX DB 3188 CTGGATCCTTGGCCCGCCACCCCTGAGTTC 3215
XX
XX RESULT 8
XX ACH04308/c
XX ID ACH04308 standard; cDNA; 424 BP.
XX AC ACH04308;
XX
XX DT 26-SEP-2003 (first entry)
XX
XX DE Human cDNA differentially expressed in lung cancer #513.
XX
XX KW Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
XX respiratory disorder; lung cancer; asthma; human.
XX
XX OS Homo sapiens.
XX
XX PN US2003065157-A1.
XX

PD 03-APR-2003.
 XX
 PP 04-APR-2002; 2002US-00116802.
 XX
 XX 04-APR-2001; 2001US-0281593P.
 PR
 XX (LASE/) LASEK A W.
 PA
 PI Lasek AW;
 XX
 XX WPI; 2003-540803/51.
 DR
 XX New combination comprising cDNAs that are differentially expressed in
 PT respiratory disorders, useful for diagnosing or treating respiratory
 PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
 PT emphysema or asthma.
 XX
 PS Claim 1; Page; 39pp; English.
 XX
 CC The invention relates to a combination comprising cDNAs or their
 CC complements that are differentially expressed in respiratory disorder.
 CC The combination is useful for preparing a composition for diagnosing or
 CC treating respiratory disorders e.g. lung cancer, chronic obstructive
 CC pulmonary disease, emphysema or asthma. The present sequence represents
 CC human cDNA differentially expressed during lung cancer
 XX
 SQ Sequence 424 BP; 73 A; 149 C; 132 G; 70 T; 0 U; 0 Other;
 Query Match 68.0%; Score 20.4; DB 9; Length 424;
 Best Local Similarity 95.5%; Pred. No. 1e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 7 CCTTGGCCCGCCCTTGAATTC 28
 DB 343 CCGTGCCCGCCCTTGAATTC 322
 RESULT 9
 ABZ18450/c
 ID ABZ18450 standard; cDNA; 466 BP.
 AC
 AC ABZ18450;
 DT 23-JAN-2003 (first entry)
 XX
 DE Group III cDNA cancer related clone SEQ ID NO:876.
 XX
 KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
 KW immune response; virology; immunology; microbiology; molecular biology;
 KW recombinant DNA technology; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200278516-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US010421.
 XX
 PR 30-MAR-2001; 2001US-0280255P.
 PR 28-AUG-2001; 2001US-0315563P.
 PR 09-JAN-2002; 2002US-0347313P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Wang T, Wang S, Bangur CS, Gaiger A;
 PI
 XX WPI; 2003-058387/05.
 DR
 XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
 PT preventing and treating cancer expressing CT or CP mRNA antigens, and in
 PT virology, immunology, microbiology, molecular biology and recombinant DNA
 PT techniques.

XX
 PS Claim 1; SEQ ID NO 876; 207pp; English.
 XX
 CC ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
 CC ABP54446 to ABP54472 represent protein (II) sequences, from the present
 CC invention. (I) and (II) have cytotatic activity and can be used in gene
 CC therapy and vaccines. (I), (II), antibodies and compositions from the
 CC present invention are useful for diagnosing, preventing and treating
 CC cancer, which expresses CT or CP mRNA antigens. They are useful for
 CC stimulating immune response. They can also be useful in virology,
 CC immunology, microbiology, molecular biology and recombinant DNA
 CC techniques. N.B. The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 466 BP; 78 A; 165 C; 144 G; 79 T; 0 U; 0 Other;
 Query Match 68.0%; Score 20.4; DB 8; Length 466;
 Best Local Similarity 95.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 7 CCTTGGCCCGCCCTTGAATTC 28
 DB 343 CCGTGCCCGCCCTTGAATTC 322
 RESULT 10
 ADD34027/c
 ID ADD34027 standard; DNA; 504 BP.
 XX
 AC ADD34027;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Mouse mitochondrial DNA sequence SEQ ID NO:1803.
 XX
 KW ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;
 KW mitochondrial disease; oxidative phosphorylation dysfunction;
 KW oxidative stress; apoptosis; aging.
 XX
 OS Mus musculus.
 XX
 PN WO2003020220-A2.
 XX
 PD 13-MAR-2003.
 XX
 PF 30-AUG-2002; 2002WO-US027886.
 XX
 PR 30-AUG-2001; 2001US-0316323P.
 PR 31-AUG-2001; 2001CA-02356540.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 XX Wallace DC, Levy S, Kerstann K, Procaccio V;
 PI
 XX WPI; 2003-300821/29.
 DR
 XX
 XX Array containing probes for genes involved in mitochondrial biology,
 PT useful for determining mitochondrial biology gene expression profiles for
 PT use in diagnosing pathologies and identifying biochemical pathways.
 XX
 PS Claim 2; SEQ ID NO 1803; 201pp; English.
 XX
 CC The invention relates to a novel array comprising at least two isolated
 CC nucleotide molecules, each molecule having a sequence capable of uniquely
 CC hybridising to a nucleic acid molecule which is an expression product of
 CC a gene involved in mitochondrial biology. The array comprises two or more
 CC isolated nucleic acid molecules or spots, each molecule having a sequence
 CC chosen from sequence of 994 human probes and 2046 mouse probes. An array
 CC of the invention is useful for determining an expression profile of a
 CC mouse or human sample containing nucleic acid, by contacting the array
 CC with the sample under conditions allowing selective hybridisation, and
 CC measuring hybridisation of nucleic acid in the sample to the array to

CC produce an expression profile. The array is also useful for determining
CC an expression profile of a first labelled sample containing nucleic acid
CC relative to a second, differently labelled sample containing nucleic
CC acid. The second sample is a reference or a standard. An array is useful
CC for determining an expression profile diagnostic of an energy-metabolism-
CC related physiological condition. An array of the invention is useful for
CC determining mitochondrial biology gene expression profiles of organisms,
CC such as human, mice and closely related species, tissue and organs of
CC such organisms, which are useful for determining expression profiles
CC diagnostic of energy metabolism-related physiological conditions,
CC diagnosing such physiological conditions, identifying biochemical
CC pathways, genes, and mutations involved in such physiological conditions,
CC identifying therapeutic agents useful for preventing and/or treating such
CC physiological conditions, evaluating and/or monitoring the efficacy of
CC such therapies, and creating and identifying animal models of human
CC energy metabolism-related physiological conditions. An array is also
CC useful for defining expression signatures or profiles for mitochondrial
CC diseases, as well as distinguishing clinical disorders that result from
CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC apoptosis and aging. An array of the invention contains probes of genes
CC not previously recognised to participate in mitochondrial biology. The
CC sequences shown in ADD33224-ADD33260 represent murine mitochondrial DNA
CC clones used to make the probes of the invention. Some sequences are not
CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC 1906, 2408 and 2643.

XX Sequence 504 BP; 114 A; 121 C; 148 G; 121 T; 0 U; 0 Other;

Query Match 66.7%; Score 20; DB 10; Length 504;

Best Local Similarity 82.1%; Pred. No. 1.6e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGATCTCTTGGCCCGCCCTTGAATTC 28

|||||

490 CTGATCTCTTGGCCCAACCTGATTC 463

RESULT 11

ABT42112/c

ID ABT42112 standard; DNA; 1760 BP.

XX AC ABT42112;

XX DT 26-JUN-2003 (first entry)

XX Toxicity modelling related rat gene SEQ ID No 1814.

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;

PR 21-FEB-2002; 2002US-0357844P.

PR 15-MAR-2002; 2002US-0364134P.

PR 08-APR-2002; 2002US-0370144P.

PR 08-APR-2002; 2002US-0370206P.

PR 08-APR-2002; 2002US-0370247P.

PR 17-APR-2002; 2002US-0372794P.

PR 21-APR-2002; 2002US-0371679P.

XX (GENE-) GENE LOGIC INC.

PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI; 2003-148464/14.

XX Predicting at least one toxic effect of a compound, useful for toxicity

XX modelling, comprises preparing a gene expression profile of a tissue or

XX cell sample exposed to the compound, and comparing the gene expression

XX profile to a database.

XX Example 4; Page; 446pp; English.

XX The invention relates to a novel method of predicting at least one toxic

XX effect of a compound. The method comprises a gene expression profile of a

XX tissue or cell sample exposed to the compound, and comparing the gene

XX expression profile to a database comprising at least part of the data or

XX information given in the specification. The methods are useful for

XX predicting at least one toxic effect of a compound, predicting the renal

XX toxicity of a compound, or identifying toxicity markers in tissues or

XX cells exposed to known renal toxin. The genes are useful as toxicity

XX markers in drug screening and toxicity assays, in monitoring disease or

XX physiological states, or disease progression. This polynucleotide

XX represents a rat DNA sequence relating to the toxic effect database

XX described in the specification. NOTE: The sequence data for this patent

XX did not form part of the printed specification, but was obtained in

XX electronic format directly from the World Intellectual Property

XX Organization

XX Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;

Query Match 66.7%; Score 20; DB 10; Length 1760;

Best Local Similarity 82.1%; Pred. No. 1.7e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TGGATCTCTTGGCCCGCCCTTGAATTC 29

|||||

431 TGGGTGCTCGCCCGCCCACTCGAATTC 404

RESULT 12

AAV20872

ID AAV20872 standard; cDNA; 3374 BP.

XX AC AAV20872;

XX DT 16-JUL-1998 (first entry)

XX Murine phospholipase D 2 (mPLD2) encoding cDNA.

XX Phospholipase D; PLD; murine; PLD mediator; autoimmune disease;

XX Rheumatoid arthritis; psoriasis; ulcerative colitis; cancer; mPLD2;

XX inflammatory disease; wound healing; ds.

XX Mus sp.

XX Key Location/Qualifiers

XX CDS 139..2937

XX /*tag= a

XX /product= "murine PLD2 protein"

XX WO9810076-A2.

XX 21-FEB-2002; 2002US-0357842P.

XX 21-FEB-2002; 2002US-0357843P.

XX 21-FEB-2002; 2002US-0357844P.

XX 21-FEB-2002; 2002US-0357845P.

XX 21-FEB-2002; 2002US-0357846P.

XX 21-FEB-2002; 2002US-0357847P.

XX 21-FEB-2002; 2002US-0357848P.

XX 21-FEB-2002; 2002US-0357849P.

XX 21-FEB-2002; 2002US-0357850P.

XX 21-FEB-2002; 2002US-0357851P.

XX 21-FEB-2002; 2002US-0357852P.

XX 21-FEB-2002; 2002US-0357853P.

XX 21-FEB-2002; 2002US-0357854P.

XX 21-FEB-2002; 2002US-0357855P.

XX 21-FEB-2002; 2002US-0357856P.

XX 21-FEB-2002; 2002US-0357857P.

XX 21-FEB-2002; 2002US-0357858P.

XX 21-FEB-2002; 2002US-0357859P.

XX 21-FEB-2002; 2002US-0357860P.

XX 21-FEB-2002; 2002US-0357861P.

XX 21-FEB-2002; 2002US-0357862P.

XX 06-AUG-1997; 97WO-US013919.
XX 05-SEP-1996; 96US-0025469P.
XX (ONYX-) ONYX PHARM INC.
XX PI Frohman MA, Morris AJ;
XX WPI; 1998-193626/17.
XX P-PSDB; AAW53281.
XX DNA sequence encoding phospholipase D - useful to identify modulators to
XX treat auto-immune and inflammatory diseases.
XX Claim 5; Page 62-66; 70pp; English.
XX This cDNA encodes a murine phospholipase D (PLD) 2 (mPLD2) protein. This
XX can be used in a method for identifying mediators of PLD, which comprises
XX transfecting a cell line with an expression vector comprising nucleic
XX acid sequences encoding a PLD protein and culturing the cell line in
XX culture medium, where the PLD protein is expressed stably. An effective
XX amount of a compound sufficient to cause a detectable loss in the
XX catalytic activity of PLD is added to the culture medium, and the loss in
XX catalytic activity is detected. The PLD proteins are enzymes having
XX phosphatidylcholine specific PLD activity. The PLD polypeptides which are
XX perinuclear membrane associated require PI(4,5)P2 for in vitro activity
XX and is activated by at least 1 G protein. The PLD polypeptides that are
XX plasma membrane associated, activates cytoskeletal reorganisation
XX pathways, require PI(4,5)P2 for in vitro activity and does not require
XX Rac1m cdc42, RhoA, PKC or ARF1 for activation. A PLD mediator can be used
XX to treat autoimmune or inflammatory diseases, specifically rheumatoid
XX arthritis, psoriasis and ulcerative colitis. The mediator can also be
XX used in wound healing and for treating cancer and other diseases
XX characterised by pathogenic mitogenicity
XX Sequence 3374 BP; 753 A; 955 C; 905 G; 760 T; 0 U; 1 Other;
Query Match 66.7%; Score 20; DB 2; Length 3374;
Best Local Similarity 82.1%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
Db 2986 CTGGATCCTTGGCCCAACCTGAGTTC 3013
RESULT 13
ID ABT09364/c
XX ABT09364 standard; DNA; 655 BP.
XX AC ABT09364;
XX DT 05-DEC-2002 (first entry)
XX DE Phase-1 Rat CT gene SEQ ID No 452.
XX KW Rat; toxicity study; rat toxic response gene; toxicological response;
XX drug development; phase-1 rat CT gene; ds.
XX OS Rattus sp.
XX FN WO200266682-A2.
XX PD 29-AUG-2002.
XX PF 29-JAN-2002; 2002WO-US002935.
XX PR 29-JAN-2001; 2001US-0264933P.
XX PR 26-JUL-2001; 2001US-0308161P.
XX PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX

PI Farris G, Hicken SH, Farr SB;
XX WPI; 2002-674961/72.
XX Evaluating the toxicity of an agent, useful in drug development or in
XX determining toxicological responses to a new drug, by determining the
XX expression of rat toxicologically relevant genes in the test animal in
XX response to the test agent.
XX Disclosure; Page 218; 388pp; English.
XX The invention relates to a method used for evaluating the toxicity of an
XX agent comprising determining the expression of a rat toxic response
XX gene(s) in the test animal in response to the agent. The method is useful
XX in drug development, particularly for conducting toxicity studies and
XX analysis before a new drug or compound is approved for human consumption
XX or use. The method is also useful in determining toxicological responses
XX to a new drug. This polynucleotide sequence represents a phase-1 rat CT
XX gene of the invention
SQ Sequence 655 BP; 174 A; 147 C; 177 G; 157 T; 0 U; 0 Other;
Query Match 65.3%; Score 19.6; DB 6; Length 655;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 GATCCTTGGCCCGCCCTTGAATTC 29
Db 493 GAACCATGACCCGCTTGAATTC 468
RESULT 14
ID ADG45424/c
XX ADG45424 standard; DNA; 655 BP.
XX AC ADG45424;
XX DT 26-FEB-2004 (first entry)
XX DE Liver inflammatory predictive gene related DNA sequence.
XX KW liver; liver toxicity; liver toxicity predictive gene;
XX liver inflammation predictive gene; inflammation; gene; ds; rat.
XX OS Rattus norvegicus.
XX FN WO2003095624-A2.
XX PD 20-NOV-2003.
XX PF 09-MAY-2003; 2003WO-US014832.
XX PR 10-MAY-2002; 2002US-0379831P.
XX PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX Kier L, Nolan TD, Sankar U, Derbel M;
XX WPI; 2004-042456/04.
XX Predicting the liver toxicity in an individual to an agent by measuring
XX the expression of one or more liver toxicity predictive genes in the
XX sample from the individual treated with the agent.
XX Claim 7; Fig 4; 336pp; English.
XX The present invention describes a method for predicting the liver
XX toxicity in an individual to an agent. The method comprises: (a)
XX obtaining a biological sample from the individual treated with the agent;
XX (b) measuring the expression of one or more liver toxicity predictive
XX genes in the sample; and (c) using the test expression profile with a set
XX of reference expression profiles in a Predictive Model to determine
XX whether the agent will induce liver toxicity in the individual. Also

described: (1) predicting the liver toxicity of an agent using an in vitro system; (2) predicting the liver toxicity in a biological sample from an individual, in vitro cell cultures or explants to an agent via a programmable machine; (3) a computer program product for enabling a computer to perform Predictive Model analysis for liver toxicity on a biological sample from an individual, in vitro cell cultures or explants to an agent; (4) a computer system adopted to predict liver toxicity in a biological system from an individual, in vitro cell cultures or explants to an agent; (5) a computer program product for predicting liver toxicity from a test sample expression profile; (6) mining genes predictive of liver toxicity; (7) an integrated system for predicting liver toxicity; and (8) identifying one or more liver inflammation predictive genes. The method is useful for predicting the liver toxicity in an individual to an agent. The present sequence represents a liver toxicity predictive gene sequence that represents a 24 hour combo 3 gene, which is specifically claimed in the exemplification of the present invention.

Sequence 655 BP; 174 A; 147 C; 177 G; 157 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 12; Length 655;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GATCTTGCCCGCCCTTGAATTC 29

DB 493 GAACCATGACCGCTTGAATTC 468

RESULT 15

AAH98347/c

ID AAH98347 standard; cDNA; 4195 BP.

AC AAH98347;

DT 12-OCT-2001 (first entry)

DE Human EST-derived coding sequence SEQ ID NO: 204.

Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse; tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

diagnostics; forensic test; gene mapping; genetic disorder; biodiversity; gene therapy; nutrition; ss.

Homo sapiens.

WO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002687.

25-JAN-2000; 2000US-00491404.

17-JUL-2000; 2000US-00617746.

03-AUG-2000; 2000US-00631451.

15-SEP-2000; 2000US-00663870.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
Cao Y, Drmanac RA, Zhang J, Werhman T;

WPI; 2001-476164/51.

P-PSDB; AAM23688.

Isolated polypeptide for treatment of diseases, diagnostics, raising

antibodies and research use.

Claim 1; Page 318-319; 1275pp; English.

The present invention provides the protein and coding sequences of novel

proteins from a variety of organisms, including human, dog, cat, horse,

cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

urchin and tomato. These were derived from expressed sequence tags (ESTs)

from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a cDNA of the invention

Sequence 4195 BP; 1273 A; 753 C; 935 G; 1234 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 4195;
Best Local Similarity 84.6%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATCTTGCCCGCCCTTGAATTC 30

DB 94 ACCCTTGCCCAACCTTGAATTC 69

RESULT 16

AAS22674/c

ID AAS22674 standard; cDNA; 4385 BP.

AC AAS22674;

DT 24-OCT-2001 (first entry)

DE Human cDNA encoding a novel human protein #240.

Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory; immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;

anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002623.

25-JAN-2000; 2000US-00491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

P-PSDB; AAU14369.

Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage.

Claim 1; Page 669-670; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides.

Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour. Polypeptides of the invention can be used to raise antibodies/elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands.

Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting

CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC encodes a protein of the invention
 XX
 SQ Sequence 4385 BP; 1331 A; 801 C; 987 G; 1266 T; 0 U; 0 Other;
 Query Match 65.3%; Score 19.6; DB 4; Length 4385;
 Best Local Similarity 84.6%; Pred. No. 2.7e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 ATCCTTGCCCGCCCTTGAATTCCC 30
 DB 94 ACCCTTGCCCAACCTTGAATTCCC 69
 RESULT 17
 AAC77106/c
 ID AAC77106 standard; cDNA; 6324 BP.
 XX
 AC AAC77106;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF2661 polynucleotide sequence SEQ ID NO:5321.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; cancer; proliferative disorder; hypertension;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertensive;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB42897.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 4493-4497; 5507pp; English.
 XX
 PS AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 6324 BP; 2040 A; 1033 C; 1359 G; 1891 T; 0 U; 1 Other;
 Query Match 65.3%; Score 19.6; DB 3; Length 6324;
 Best Local Similarity 84.6%; Pred. No. 2.8e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 ATCCTTGCCCGCCCTTGAATTCCC 30
 DB 4773 ACCCTTGCCCAACCTTGAATTCCC 4748
 RESULT 18
 ADB16929/c
 ID ADB16929 standard; DNA; 50000 BP.
 XX
 AC ADB16929;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human DYXC1 DNA, chromosomal gene region nucleotides 150001-200000.
 XX
 KW ds; human; DYXC1; dyslexia; neurological disorder; chromosome 15q21;
 KW reading disability; phonological processing; rapid naming;
 KW verbal short-term memory.
 XX
 OS Homo sapiens.
 XX
 PN WO20003068814-A1.
 XX
 PD 21-AUG-2003.
 XX
 PF 12-FEB-2003; 2003WO-FI000110.
 XX
 PR 12-FEB-2002; 2002US-0355782P.
 XX
 PA (LICN) LICENTIA LTD.
 XX
 PI Kere J, Taipale M, Nopola-Hemmi J, Kaminen N;
 XX
 DR WPI; 2003-646482/61.
 XX
 PT New isolated, purified DYXC1 nucleic acid for studying brain processes,
 PT e.g. reading, phonological processing, rapid naming or verbal short-term
 PT memory, or for diagnosing dyslexia or assessing the predisposition to
 PT dyslexia.
 XX
 PS Claim 27; Page 97-110; 135pp; English.
 XX
 CC This invention relates to a novel isolated human gene DYXC1 that is
 CC functionally related to dyslexia, more particularly it describes single
 CC nucleotide polymorphisms thought to predispose an individual in to
 CC developing dyslexia. This is a neurological disorder with a genetic basis
 CC (DYXC1 has been isolated to chromosome 15q21), which manifests itself as
 CC a specific reading disability. Specifically, DYXC1 is can be useful in

CC study of brain processes such as reading, phonological processing, rapid
 CC naming and verbal short-term memory. Accordingly, the present invention
 CC describes methods and materials for analysing allelic variations in the
 CC DYX1C1 gene, and also provides DYX1C1 as an antigen for the production of
 CC antibodies used in the diagnosis of dyslexia. This polynucleotide is the
 CC partial genomic sequence of the human DYX1C1 chromosomal region
 CC (nucleotides 150001-200000) of the invention.

XX SQ Sequence 50000 BP; 14103 A; 9577 C; 10417 G; 15903 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 9; Length 50000;
 Best Local Similarity 84.6%; Pred. No. 3.3e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATCTTGGCCCGCCCTTGAATTC 30
 DB 42561 ACCCTTGGCCCAACCTTGAATTC 42536

RESULT 19
 ID AAQ61741/c
 AC AAQ61741 standard; cDNA; 33 BP.

XX AC AAQ61741;

DT 25-MAR-2003 (revised)
 DT 21-OCT-1994 (first entry)

XX DE HEV strain BUR-121 primer R112.

XX KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF; PCR;
 KW antibody; detection; diagnosis; primates; stool suspension; amplify;
 KW polymerase chain reaction; primer; burma; strain BUR-121; ss.

XX OS Synthetic.

XX PN WO9406913-A2.

XX PD 31-MAR-1994.

XX PF 17-SEP-1993; 93WO-US008849.

XX PR 18-SEP-1992; 92US-00947263.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Tsarev SA, Emerson SU, Purcell RH;

XX DR WPI; 1994-118462/14.

XX PT Purified hepatitis E strain SAR-55 virus - used to develop prods. for use
 PT in detection, diagnosis, vaccines and therapy of hepatitis E virus
 PT infection.

XX PS Example 1; Page 39; 114pp; English.

XX CC The sequences given in AAQ45198-200 and AAQ61687-777 are primers which
 CC were used in the isolation and amplification of the genomic sequence of
 CC the hepatitis E virus (HEV) strain SAR-55. These primers were based on
 CC sequences derived from the SAR-55 strain and a strain from Burma (BUR-
 CC 121). The amplified sequence contains three open reading frames (ORFs).
 CC The proteins encoded by this sequence can be used to stimulate the
 CC production of protective antibodies upon injection into a mammal that
 CC would serve to protect the mammal upon challenge with wild type HEV. The
 CC proteins can be used for detection and diagnosis of HEV infection. This
 CC cDNA was isolated from primates inoculated with stool suspensions
 CC obtained from hepatitis E patients. (Updated on 25-MAR-2003 to correct PN
 CC field.)

XX SQ Sequence 33 BP; 11 A; 6 C; 13 G; 3 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 2; Length 33;
 Best Local Similarity 79.3%; Pred. No. 2.2e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
 DB 32 CTGGTCTTTATGCGCGCCCTCGAATTC 4

RESULT 20
 ID AAT27452/c
 AC AAT27452 standard; DNA; 33 BP.

XX AC AAT27452;

XX DT 27-NOV-1996 (first entry)

XX DE HEV strain Burma-121 derived reverse primer 112 (ORF-2).

XX KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
 KW structural region; antigen; detection; antibody; vaccine; immunisation;
 KW infection; primer; Burma-121; polymerase chain reaction; PCR; ss.

XX OS Synthetic.

XX PN WO9610580-A2.

XX PD 11-APR-1996.

XX PF 03-OCT-1995; 95WO-US013102.

XX PR 03-OCT-1994; 94US-00316765.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Tsarev SA, Emerson SU, Purcell RH;

XX DR WPI; 1996-209320/21.

XX PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
 PT antigenic protein useful in diagnosis, prophylaxis and treatment of
 PT hepatitis E virus infection.

XX PS Example 1; Page 42; 121pp; English.

XX CC The present sequence is a hepatitis E virus (HEV) strain Burma-121
 CC derived primer, used in the isolation of the HEV strain SAR-55 cDNA. The
 CC HEV strain SAR-55 was implicated in an enterically transmitted non-A, non
 CC -B hepatitis in Pakistan. The protein encoded by the structural region of
 CC the virus (i.e. ORF-2), which is capable of forming HEV like particles,
 CC is useful for the detection of HEV antibodies (pref. IgG or IgM) in
 CC blood, plasma, sera, cerebrospinal fluid, tissue, urine or pleural fluid.
 CC The protein, and anti-HEV antibodies generated using the protein, can
 CC also be used in vaccines for immunising an animal against HEV infection.
 CC The protein is identified as a band of greater than 50 kD following SDS-
 CC PAGE of cell lysates of insect cells infected with a HEV ORF-2 con9.
 CC baculovirus, i.e. the claimed recombinant expression vectors pPIC9-1779,
 CC -1780 and -1781

XX SQ Sequence 33 BP; 11 A; 6 C; 13 G; 3 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 2; Length 33;
 Best Local Similarity 79.3%; Pred. No. 2.2e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
 DB 32 CTGGTCTTTATGCGCGCCCTCGAATTC 4

RESULT 21
 ID AAV71662/c
 AC AAV71662 standard; DNA; 33 BP.

XX AC AAV71662;

```

XX 02-FEB-1999 (first entry)
XX HEV ORF proteins encoding DNA amplifying primer R 112 B.
DE Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF; PCR primer; ss.
XX Synthetic.
OS Hepatitis E virus.
XX WO9846761-A1.
PN 22-OCT-1998.
XX 09-APR-1998; 98WO-US007418.
XX 11-APR-1997; 97US-00840316.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA Emerson SU, Purcell RH, Tsarev SA, Robinson RA;
PI WPI; 1998-568733/48.
XX New hepatitis E virus DNA from Pakistani strain SAR-55 - used for, e.g.
DR developing products for diagnosis of, and vaccination against hepatitis E
PT virus infection.
PT virus infection.
XX Example 1; Page 44; 204pp; English.
XX Sequences AAV71605 to AAV71698 represent primers used for PCR
CC amplification of the hepatitis E virus (HEV) DNA SAR-55 encoding the open
CC reading frame (ORF) proteins ORF-1, ORF-2 and ORF-3. A host organism
CC transformed or transfected with a recombinant expression vector
CC containing the SAR-55 nucleic acid can be used to produce the HEV
CC proteins, especially ORF-2 protein. The recombinant HEV proteins can be
CC used as diagnostic agents and as vaccines for use against HEV infection.
CC The detection of antibodies specific for HEV can be used for the
CC diagnosis of infection and diseases caused by HEV, and for monitoring the
CC progression of such disease. Such methods are also useful for monitoring
CC the efficacy of therapeutic agents during the course of treatment of HEV
CC infection and disease in a mammal. The antibodies can be used for
CC detection or for passive immunisation of mammals
XX Sequence 33 BP; 11 A; 6 C; 13 G; 3 T; 0 U; 0 Other;
SQ
Query Match 64.7%; Score 19.4; DB 2; Length 33;
Best Local Similarity 79.3%; Pred. No. 2.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 32 CTGGTCTTTATGCGCCCGCCCTGGAATTC 4
RESULT 22
ADL85957
ID ADL85957 standard; DNA; 540 BP.
XX ADL85957;
AC ADL85957;
XX 20-MAY-2004 (first entry)
XX DNA up-regulated in murine multipotent progenitor cells SeqID 2350.
DE gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX Mus sp.
OS Mus sp.
XX WO2003093445-A2.
PN

```

```

XX 13-NOV-2003.
XX 05-MAY-2003; 2003WO-US014114.
XX 03-MAY-2002; 2002US-0377383P.
XX (STOW-) STOWERS INST MEDICAL RES.
XX Li L;
XX WPI; 2004-022656/02.
XX Classifying an unknown multi-lineage affiliated gene comprises isolating
PT expressed nucleic acid sequences from the discrete cell sub-populations.
PT
XX Claim 8; SEQ ID NO 2350; 123pp; English.
XX This invention relates to a novel method for predicting gene potential by
CC associating nucleic acid sequences of unknown function with particular
CC sub-population profiles. Specifically, it refers to classifying an
CC unknown multi-lineage affiliated gene by collecting hybridisation data to
CC develop a gene expression map, in order to determine the discrete sub-
CC population where it is expressed. The present invention describes methods
CC for predicting the lineage commitment of genes associated with the self-
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
CC referred to as bone marrow stem cells populations. As such, these methods
CC can be used to identify associated multi-lineage affiliated genes and
CC hence the underlying molecular mechanisms in physiological haematopoietic
CC development. This polynucleotide sequence is DNA associated with a murine
CC MPP sub population of cells of the invention.
XX Sequence 540 BP; 93 A; 146 C; 157 G; 128 T; 0 U; 16 Other;
SQ
Query Match 64.7%; Score 19.4; DB 12; Length 540;
Best Local Similarity 79.3%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 276 TGCTTCTTCTTATGCGCCCGCCCTGGAATTCAC 304
RESULT 23
ADL85956
ID ADL85956 standard; DNA; 540 BP.
XX ADL85956;
AC ADL85956;
XX 20-MAY-2004 (first entry)
XX DNA up-regulated in murine multipotent progenitor cells SeqID 2349.
DE gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX Mus sp.
OS Mus sp.
XX WO2003093445-A2.
PN
XX 13-NOV-2003.
XX 05-MAY-2003; 2003WO-US014114.
XX 03-MAY-2002; 2002US-0377383P.
XX (STOW-) STOWERS INST MEDICAL RES.
XX Li L;
XX

```

DR WPI; 2004-022656/02.
XX Classifying an unknown multi-lineage affiliated gene comprises isolating
PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX
XX Claim 8; SEQ ID NO 2349; 123pp; English.
XX
XX This invention relates to a novel method for predicting gene potential by
CC associating nucleic acid sequences of unknown function with particular
CC sub-population profiles. Specifically, it refers to classifying an
CC unknown multi-lineage affiliated gene by collecting hybridisation data to
CC develop a gene expression map, in order to determine the discrete sub-
CC population where it is expressed. The present invention describes methods
CC for predicting the lineage commitment of genes associated with the self-
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
CC referred to as bone marrow stem cells populations. As such, these methods
CC can be used to identify associated multi-lineage affiliated genes and
CC hence the underlying molecular mechanisms in physiological haematopoietic
CC development. This polynucleotide sequence is DNA associated with a murine
CC MPP sub population of cells of the invention.
XX
XX Sequence 540 BP; 93 A; 146 C; 157 G; 128 T; 0 U; 16 Other;
Query Match 64.7%; Score 19.4; DB 12; Length 540;
Best Local Similarity 79.3%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 276 TGCTTCCTCCCTGCCCTTGAATCCAC 304
RESULT 24.
ACC61545/C
ID ACC61545 standard; DNA; 1041 BP.
XX
AC ACC61545;
XX
DT 20-JUN-2003 (first entry)
XX
DE Gene sequence #SEQ ID 1872.
XX
XX Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PP 20-DEC-2001; 2001EP-00130253.
XX
XX 15-MAY-2001; 2001EP-00111774.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse U, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX
DR WPI; 2003-250078/25.
DR P-PSDB; ABR53503.
XX
XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
XX Disclosure; SEQ ID NO 1872; 17pp + Sequence Listing; English.
XX
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records

CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM
XX
SQ Sequence 1041 BP; 396 A; 231 C; 191 G; 223 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 10; Length 1041;
Best Local Similarity 79.3%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 1030 TGGATCCTTGGCCCTTCTCTTGAGTTTC 1002
RESULT 25
ADK64479/C
ID ADK64479 standard; DNA; 1041 BP.
XX
AC ADK64479;
XX
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived gene #1136.
XX
XX protein complex; drug target; diagnosis; gene; ds.
XX
OS Unidentified.
XX
PN EP1338608-A2.
XX
PD 27-AUG-2003.
XX
PP 20-DEC-2002; 2002EP-00102902.
XX
XX 20-DEC-2001; 2001EP-00130253.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX
DR WPI; 2003-638460/61.
DR P-PSDB; ADK64478.
XX
XX New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
XX
XX Disclosure; SEQ ID NO 2272; 13pp; English.
XX
XX The invention relates to novel protein complexes comprising a first and a
CC second protein, or its derivative, fragment, homologue or variant. The
CC proteins are selected from given protein complexes, which are not defined
CC in the specification. The variants are encoded by nucleic acids that
CC hybridize to the nucleic acids encoding the proteins under low stringency
CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein

CC components of the complex is useful for the manufacture of a medicament
 CC for the treatment or prevention of a disease or disorder. This sequence
 CC corresponds to a gene of the invention. (Note: the sequence data for this
 CC patent did not form part of the printed specification but was obtained
 CC from the EPO in electronic format).

XX SQ Sequence 1041 BP; 396 A; 231 C; 191 G; 223 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 1041;
 Best Local Similarity 79.3%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TGGATCCTTGGCCCGCCCTTGAATTCCC 30
 ||||| ||||| ||||| ||||| |||||
 Db 1030 TGGATCCTTGGCCCTTCTCTTGAGTTCTC 1002

RESULT 26

AD547118/c
 ID ADS47118 standard; cDNA; 1041 BP.

XX AC ADS47118;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polynucleotide #1861.

XX KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.

XX OS Bacteria.

XX PN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

XX PA (HINK/) HINKLE G J.

XX PA (SLAT/) SLATER S C.

XX PA (CHEN/) CHEN X.

XX PA (GOLD/) GOLDMAN B S.

XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX DR WPI; 2004-061375/06.

XX PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.

XX PS Claim 1; SEQ ID NO 25548; 122pp; English.

XX CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 1041 BP; 396 A; 231 C; 191 G; 223 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 13; Length 1041;
 Best Local Similarity 79.3%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TGGATCCTTGGCCCGCCCTTGAATTCCC 30

||||| ||||| ||||| ||||| |||||
 Db 1030 TGGATCCTTGGCCCTTCTCTTGAGTTCTC 1002

RESULT 27

ADBI1679

ID ADBI1679 standard; DNA; 1140 BP.

XX AC ADBI1679;

XX DT 20-NOV-2003 (first entry)

XX DS Alloiooccus otitis antigenic protein encoding DNA SEQ ID NO:5383.

XX KW Alloiooccus otitis; antigenic protein; immunogenic; immunisation;
 KW gene therapy; Gram-positive bacterium; infection; gene; ds.

XX OS Alloiooccus otitis.

XX PN WO2003048304-A2.

XX PD 12-JUN-2003.

XX PF 25-NOV-2002; 2002WO-US036123.

XX PR 29-NOV-2001; 2001US-0333777P.

XX PR 18-NOV-2002; 2002US-0426742P.

XX PA (AMHP) WYETH HOLDINGS CORP.

XX PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX DR WPI; 2003-505284/47.

XX DR P-PSDB; ADBI1676.

XX PT New Alloiooccus otitis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.

XX PS Claim 7; SEQ ID NO 5383; 1019pp; English.

XX CC The present invention describes an isolated polynucleotide (1) of
 CC Alloiooccus otitis genomic DNA, which encodes an antigenic protein.
 CC Alloiooccus otitis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
 CC expression vector comprising the novel isolated polynucleotide (1), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transformed, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array

CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against *Alloiooccus* otitis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*
 CC otitis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (1) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting *Alloiooccus*
 CC otitis. The present sequence encodes an *Alloiooccus* otitis antigen
 CC protein from the present invention.
 XX
 SQ Sequence 1140 BP; 319 A; 278 C; 253 G; 290 T; 0 U; 0 Other;
 Query Match 64.7%; Score 19.4; DB 9; Length 1140;
 Best Local Similarity 79.3%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCC 29
 DB 483 CTGGATCCTTGGCCCGCCCTTCAATTAC 511
 RESULT 28
 AAD05374/c
 ID AAD05374 standard; cDNA; 1396 BP.
 AC AAD05374;
 XX
 XX 17-JUL-2001 (first entry)
 DT Human secreted protein-encoding gene 24 cDNA clone HNG0112, SEQ ID NO:85.
 DE Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
 KW chemotaxis; food additive; gene therapy; binding partner identification;
 KW 86.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 596..877
 FT /*tag= a
 FT /product= "Human secreted protein"
 FT /note= "CDS does not include start codon"
 FT /transl_except= (pos:665..667, aa:Xaa)
 FT /transl_except= (pos:788..790, aa:Xaa)
 FT /note= "Xaa equals any of the twenty naturally occurring L
 FT -amino acids"
 FT /partial
 FT sig_peptide 596..670
 FT /*tag= b
 FT mat_peptide 671..874
 FT /*tag= c
 FT /product= "Human mature secreted protein"
 XX
 XX WO200134626-A1.
 PN
 XX 17-MAY-2001.
 XX
 XX 01-NOV-2000; 2000NO-US030045.
 XX
 XX 05-NOV-1999; 99US-0163581P.

PR 30-JUN-2000; 2000US-0215133P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;
 XX WPI: 2001-308778/32.
 DR P-PSDB; AAE01508.
 XX New nucleic acid molecules encoding 28 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX Claim 1; Page 478-479; 562pp; English.
 XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted
 CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.
 CC The genes and their secreted proteins are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 28 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein-encoding cDNA of the
 CC invention
 XX Sequence 1396 BP; 280 A; 446 C; 432 G; 228 T; 0 U; 10 Other;
 SQ
 Query Match 64.7%; Score 19.4; DB 4; Length 1396;
 Best Local Similarity 79.3%; Pred. No. 3e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCC 29
 DB 572 CTGGATCCTTGGCCCGGAACCTGCATTCC 544
 RESULT 29
 ACC50784/c
 ID ACC50784 standard; cDNA; 1396 BP.
 XX ACC50784;
 AC
 XX
 XX 12-JUN-2003 (first entry)
 XX Human secreted protein coding sequence, SEQ ID 451.
 DE Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
 KW vulnery; antiinflammatory; nootropic; neuroprotective;
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200295010-A2.


```

PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX
PR 19-JUL-2001; 2001US-0306171P.
PR
PR 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI; 2003-430516/40.
DR
DR P-PSDB; ADC74498.
XX
XX Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
PT
PS Claim 9; SEQ ID NO 332; 1537pp; English.
XX
XX The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB9148 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1396 BP; 280 A; 446 C; 432 G; 228 T; 0 U; 10 Other;

Query Match 64.7%; Score 19.4; DB 9; Length 1396;
Best Local Similarity 79.3%; Pred. No. 3e+02; Mismatches 0; Gaps 0;
Matches 23; Conservative 0; Indels 6;

OY 1 CTGGATCCTTGCCCGCCCGCCCTTGAAATCC 29
Db 572 CTGGATCCTTGCCCGCGAACCTGCATCC 544

RESULT 32
ADC73883/c
ID ADC73883 standard; DNA; 1396 BP.
XX
XX ADC73883;
XX
XX 01-JAN-2004 (first entry)
DT
XX
XX Human secreted protein-related DNA - SEQ ID 516.
DE
XX antianemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
KW antidiabetic; immunosuppressive; dermatological; nephrotropic;
KW antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
KW fungicide; antiparasitic; antiarteriosclerotic; vulnerary; cytostatic;
KW haemopoietic; haematologic; anaemia; autoimmune disorder;
KW rheumatoid arthritis; inflammation; Grave's disease; diabetes;
KW systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
KW Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
KW cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
KW human; gene; ds.
XX
XX Homo sapiens.
OS
XX WO2003038063-A2.
PN
XX 08-MAY-2003.
PD

19-MAR-2002; 2002WO-US008277.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX
PR 19-JUL-2001; 2001US-0306171P.
PR
PR 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI; 2003-430516/40.
DR
DR P-PSDB; ADC74498.
XX
XX New human secreted polypeptide for diagnosing, preventing or treating
PT hemopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).
PT
XX Claim 27; SEQ ID NO 516; 2272pp; English.
XX
XX The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein-related DNA of
CC the invention.
XX
SQ Sequence 1396 BP; 280 A; 446 C; 432 G; 228 T; 0 U; 10 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 1396;
Best Local Similarity 79.3%; Pred. No. 3e+02; Mismatches 0; Gaps 0;
Matches 23; Conservative 0; Indels 6;

OY 1 CTGGATCCTTGCCCGCCCGCCCTTGAAATCC 29
Db 572 CTGGATCCTTGCCCGCGAACCTGCATCC 544

RESULT 33
ADD37762/c
ID ADD37762 standard; cDNA; 1396 BP.
XX
XX ADD37762;
XX
XX 15-JAN-2004 (first entry)
DT
XX
XX Human secreted protein encoding sequence #244.
DE
XX human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
KW Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic; ss.
XX
XX Homo sapiens.
OS
XX WO200290526-A2.
PN
XX 14-NOV-2002.
PD
XX
XX 19-MAR-2002; 2002WO-US008279.
XX
XX 21-MAR-2001; 2001US-0277340P.
PR
PR 19-JUL-2001; 2001US-0306171P.
PR
PR 13-NOV-2001; 2001US-0331287P.

```

```
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX PI WPI; 2003-140218/13.
XX DR
XX PA New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing or
PT treating allergic or asthmatic disorders, or related immediate
PT hypersensitivity disorders.
XX PS Claim 7; SEQ ID NO 244; 1323pp; English.
XX CC The present invention relates to an isolated polypeptide or human
CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
CC their fragments, and agonists or antagonists that bind are useful for
CC preparing a diagnostic or pharmaceutical composition for diagnosing or
CC treating allergic or asthmatic disorders. The polypeptide is also useful
CC for identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases the activity of the polypeptide. The polypeptides and nucleic
CC acid molecules are also useful for detecting, preventing, diagnosing,
CC prognosticating, treating or ameliorating inflammatory disorders
CC neoplastic diseases, wound healing and disorders of epithelial cell
CC proliferation, immune disorders, cardiovascular disorders, blood-related
CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
CC disorders. The nucleic acids are also useful for chromosome
CC identification, radiation hybrid mapping or long-range restriction
CC mapping, as molecular weight markers, or as hybridization or diagnostic
CC probes. The polypeptides and antibodies are useful for providing
CC immunological probes for differential identification of the tissues
CC immunohistochemistry assays. The present sequence represents a human
CC secreted protein encoding sequence.
XX SQ Sequence 1396 BP; 280 A; 446 C; 432 G; 228 T; 0 U; 10 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 1396;
Best Local Similarity 79.3%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTCTGCGCCGCCCTTGAAATTC 29
Db ||||| ||||| ||||| ||||| |||||
572 CTGGATCTCTGCGCGGAACCTGCATTC 544

RESULT 34
ADA56580/c
ID ADA56580 standard; DNA; 1396 BP.
XX AC ADA56580;
XX DT 20-NOV-2003 (first entry)
XX DE Gene encoding human secreted protein #373.
XX KW immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
KW cystostatic; cerebroprotective; neuroprotective; nootropic;
KW cardiovascular; antiarteriosclerotic; gene therapy;
KW human secreted protein; immune disorder; inflammation;
KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;
KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
KW triple helix formation; antisense gene therapy; forensic biology; ds;
KW gene.
XX OS Homo sapiens.
XX PN WO2002102994-A2.
XX PD 27-DEC-2002.
XX KW
```

```
PF 19-MAR-2002; 2002WO-US008278.
XX 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI; 2003-167512/16.
XX P-PSDB; ADA57473.
XX New human secreted polypeptides and polynucleotides, useful for
PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
PT conditions, respiratory disorders, cancers, CNS disorders, or
PT neurodegenerative disorders.
XX Claim 21; SEQ ID NO 769; 1754pp; English.
XX The invention relates to 592 new human secreted polypeptides useful for
CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
CC conditions, respiratory disorders, cancers, CNS disorders, or
CC neurodegenerative disorders, or polypeptides comprising an amino acid
CC sequence at least 93% identical to the new sequences. The polypeptides,
CC antibodies or antibody fragments that bind to the polypeptides, nucleic
CC acids encoding the polypeptides, agonists or antagonists that binds to
CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
CC compositions for diagnosing, treating or preventing an e.g. immune
CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
CC disorders (e.g. Parkinson's disease or Alzheimer's disease) and
CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
CC polynucleotides are useful for chromosome identification, chromosome
CC mapping, for controlling gene expression through triple helix formation
CC or antisense DNA or RNA, in gene therapy, for identifying individuals
CC from minute biological samples, in forensic biology, and as hybridization
CC probes. The polypeptides are useful for as molecular weight markers on
CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
CC gels, to raise antibodies, for testing biological activities, and for
CC treating or preventing neural disorders, immune system disorders,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, proliferative and/or cancerous diseases. This sequence corresponds
CC to a gene encoding one of the polypeptide of the invention. Note: The
CC sequence data for this patent did form part of the printed specification,
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1396 BP; 280 A; 446 C; 432 G; 228 T; 0 U; 10 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 1396;
Best Local Similarity 79.3%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTCTGCGCCGCCCTTGAAATTC 29
Db ||||| ||||| ||||| ||||| |||||
572 CTGGATCTCTGCGCGGAACCTGCATTC 544

RESULT 35
AAD05323
ID AAD05323 standard; cDNA; 2128 BP.
XX AC AAD05323;
XX DT 17-JUL-2001 (first entry)
XX DE Human secreted protein-encoding gene 24 cDNA clone HNG0112, SEQ ID NO:34.
XX KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
```

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiodysplasia; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulvovaginitis; cell culture;
KW chemotaxis; food additive; gene therapy; binding partner identification;
KW BB.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 27..200

XX /tag= a

XX /product= "Human secreted protein"

XX sig_peptide 27..104

XX /tag= b

XX mat_peptide 105..197

XX /tag= c

XX /product= "Human mature secreted protein"

XX WO200134626-A1.

XX 17-MAY-2001.

XX 01-NOV-2000; 2000WO-US030045.

XX 05-NOV-1999; 99US-0163581P.

XX 30-JUN-2000; 2000US-0215133P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;

XX WPI; 2001-308778/32.

XX P-PSDB; AA001459.

XX New nucleic acid molecules encoding 28 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.

XX Claim 1; Page 440-441; 562pp; English.

XX AAD05300-AA05379 represent cDNAs corresponding to 28 human secreted
CC protein genes, and AA051436-AA051513 represent the proteins they encode.
CC AA051514-AA051544 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 28 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiodysplasia, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC present sequence represents a human secreted protein-encoding cDNA of the
CC invention

XX SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

XX Query Match 64.7%; Score 19.4; DB 4; Length 2128;

XX Best Local Similarity 79.3%; Pred. No. 3.1e+02;

XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29

DB 1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585

RESULT 36

ACC50558

ID ACC50558 standard; cDNA; 2128 BP.

XX ACC50558;

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 225.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;

XX vulvovaginitis; antiinflammatory; nootropic; neuroprotective;

XX antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.

XX Homo sapiens.

XX WO200295010-A2.

XX 28-NOV-2002.

XX 19-MAR-2002; 2002WO-US009785.

XX 21-MAR-2001; 2001US-0277340P.

XX 19-JUL-2001; 2001US-0306171P.

XX 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-129429/12.

XX Novel human secreted proteins, useful for detecting, preventing,

XX diagnosing, prognosticating, treating and/or ameliorating cardiovascular

XX disorders such as arrhythmia.

XX Claim 21; SEQ ID NO 225; 1881pp; English.

XX The present invention relates to novel human secreted proteins (ABR47633-

XX ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins

XX and their coding sequences are useful for the preparation of a diagnostic

XX or pharmaceutical composition for diagnosing or treating a cardiovascular

XX disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary

XX arteriosclerosis and myocardial ischaemia), neural disorders, immune

XX system disorders, muscular disorders, reproductive disorders,

XX gastrointestinal disorders, pulmonary disorders, renal disorders,

XX proliferative disorders and/or cancerous diseases and conditions, for

XX wound healing and epithelial cell proliferation, to treat inflammation or

XX infection, for treating thrombosis and arteriosclerosis, for treating or

XX preventing neural damage which occurs in neuronal disorders or

XX neurodegenerative conditions such as Alzheimer's disease and Parkinson's

XX disease, to enhance bone and periodontal regeneration and aid in tissue

XX transplants or bone grafts, to prevent skin aging or hair loss, to

XX stimulate growth and differentiation of haematopoietic cells and bone

XX marrow cells when used in combination with other cytokines, to maintain

XX organs before transplantation or for supporting cell culture of primary

XX tissues, to increase or decrease differentiation or proliferation of

XX embryonic stem cells, or to modulate mammalian characteristics or

XX metabolism. Note: The sequence data for this patent was published in

XX electronic format and is available from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences

```
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

Query Match          64.7%; Score 19.4; DB 8; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCGCCGCCCTTGAATTC 29
   ||||| ||||| ||||| ||||| |||||
Db 1557 CTGGATCCTTGCCTCCGCCGCCCTTGAATTC 1585

RESULT 37
ACC50920
ID ACC50920 standard; cDNA; 2128 BP.
XX
AC ACC50920;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human secreted protein BAC clone SEQ ID NO 1100.
XX
KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnerary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200295010-A2.
XX
PD 28-NOV-2002.
XX
PF 19-MAR-2002; 2002WO-US009785.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
PI WPI; 2003-129429/12.
XX
PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Disclosure; SEQ ID NO 1100; 1881pp; English.
XX
CC The present invention relates to novel human secreted proteins (ABR47633-
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC infection, for treating thrombosis and arteriosclerosis, for treating or
CC preventing neural damage which occurs in neuronal disorders or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of haematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. The present sequence was used to illustrate the invention.
CC Note: The sequence data for this patent was published in electronic
CC format and is available from WIPO at
```

```
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

Query Match          64.7%; Score 19.4; DB 8; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCGCCGCCCTTGAATTC 29
   ||||| ||||| ||||| ||||| |||||
Db 1556 CTGGATCCTTGCCTCCGCCGCCCTTGAATTC 1584

RESULT 38
ABZ71307
ID ABZ71307 standard; cDNA; 2128 BP.
XX
AC ABZ71307;
XX
DT 04-APR-2003 (first entry)
XX
DE Secreted protein-encoding gene 118 cDNA clone HNG0112, SEQ ID NO:128.
XX
KW Human; secreted protein; digestive disorder; gastrointestinal disorder;
KW mouth; oesophagus; stomach; small intestine; large intestine; liver;
KW biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
KW immune disorder; inflammation; infection; wound healing; drug screening;
KW chromosome identification; chromosome mapping; cytostatic; gene therapy;
KW antiinflammatory; immunosuppressive; vulnerary; chromosome 11; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200276488-A1.
XX
PD 03-OCT-2002.
XX
PF 19-MAR-2002; 2002WO-US008276.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
PI WPI; 2003-029900/02.
XX
PI P-PSDB; ABR00128.
XX
PT New human secreted proteins and nucleic acids, useful for detecting,
PT preventing, diagnosing, prognosticating, treating and/or ameliorating
PT e.g. gastrointestinal diseases and disorders, or cancers.
XX
PS Claim 21; Page 826-827; 1216pp; English.
XX
CC ABZ71190-ABZ71478 represent cDNAs corresponding to 178 human secreted
CC protein genes, and ABP00011-ABP00299 represent the proteins they encode.
CC ABZ71479-ABZ71540 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening, and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing, treating, ameliorating or preventing
CC digestive disorders. Such conditions include disorders of the mouth,
CC oesophagus, stomach, small intestine, large intestine, liver, biliary
CC tract and pancreas, and include cancers of these organs and tissues. The
CC secreted proteins and their nucleic acids may also be used in the
CC treatment of immune disorders, inflammation, infection,
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids
CC of the invention may be used for chromosome identification, chromosome
CC mapping, in gene therapy, for identifying individuals from minute
CC biological samples, as hybridisation probes, and as molecular weight
```

CC markers. The present sequence represents a human secreted protein-
CC encoding cDNA clone of the invention
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 8; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGCCCGCCCTTGAATTC 29
Db 1557 CTGGATCCTTGCCCGCCGACCTGCATTCC 1585
RESULT 39
ABZ71522
ID ABZ71522 standard; DNA; 2128 BP.
XX
AC ABZ71522;
XX
DT 03-APR-2003 (first entry)
XX
DE Secreted protein gene 118 genomic fragment HNGO112, SEQ ID NO:632.
XX
KW Human; secreted protein; digestive disorder; gastrointestinal disorder;
KW mouth; oesophagus; stomach; small intestine; large intestine; liver;
KW biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
KW immune disorder; inflammation; infection; wound healing; drug screening;
KW chromosome identification; chromosome mapping; cytostatic; gene therapy;
KW antiinflammatory; immunosuppressive; vulnery; chromosome 11; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200276488-A1.
XX
PD 03-OCT-2002.
XX
PF 19-MAR-2002; 2002WO-US008276.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-029900/02.
XX
PT New human secreted proteins and nucleic acids, useful for detecting,
PT preventing, diagnosing, prognosticating, treating and/or ameliorating
PT e.g. gastrointestinal diseases and disorders, or cancers.
XX
PS Disclosure; Page 1190; 1216pp; English.
XX
CC ABZ71190-ABZ71478 represent cDNAs corresponding to 178 human secreted
CC protein genes, and ABP00011-ABP00299 represent the proteins they encode.
CC ABZ71479-ABZ71540 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening, and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing, treating, ameliorating or preventing
CC digestive disorders. Such conditions include disorders of the mouth,
CC oesophagus, stomach, small intestine, large intestine, liver, biliary
CC tract and pancreas, and include cancers of these organs and tissues. The
CC secreted proteins and their nucleic acids may also be used in the
CC treatment of immune disorders, inflammation, infection,
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids
CC of the invention may be used for chromosome identification, chromosome
CC mapping, in gene therapy, for identifying individuals from minute
CC biological samples, as hybridisation probes, and as molecular weight

CC markers. The present sequence represents a human secreted protein genomic
CC fragment referred to in the disclosure of the invention
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 8; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGCCCGCCCTTGAATTC 29
Db 1556 CTGGATCCTTGCCCGCCGACCTGCATTCC 1584
RESULT 40
ADB91226
ID ADB91226 standard; cDNA; 2128 BP.
XX
AC ADB91226;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human secreted protein cDNA #SEQ ID 172.
XX
KW Secreted protein; gene therapy; antidiabetic; diabetes; human; ss.
XX
OS Homo sapiens.
XX
PN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
XX
PS Claim 9; SEQ ID NO 172; 1537pp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 9; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGCCCGCCCTTGAATTC 29

```
Db      1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585
||||| ||||| ||||| ||||| |||||
RESULT 41
ADB91883
ID ADB91883 standard; DNA; 2128 BP.
AC ADB91883;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human secreted protein related DNA #SEQ ID 829.
XX
KW Secreted protein; gene therapy; antidiabetic; diabetes; human; ds.
XX
OS Homo sapiens.
XX
FN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
XX
PR 19-JUL-2001; 2001US-0306171P.
XX
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
XX
PS Disclosure; SEQ ID NO 829; 1537pp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 9; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
||||| ||||| ||||| ||||| |||||
Db      1556 CTGGATCTTGGCCCGGAACCTGCATTC 1584

RESULT 42
ADC73623
ID ADC73623 standard; DNA; 2128 BP.
XX
AC ADC73623;
XX
```

```
DT 01-JAN-2004 (first entry)
XX
DE Human secreted protein-related DNA - SEQ ID 256.
XX
KW antianemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
KW antidiabetic; immunosuppressive; dermatological; nephrotropic;
KW antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
KW fungicide; antiparasitic; antiarteriosclerotic; vulnerary; cytostatic;
KW haemopoietic; haematologic; anaemia; autoimmune disorder;
KW rheumatoid arthritis; inflammation; Grave's disease; diabetes;
KW systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
KW Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
KW cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
KW human; gene; ds.
XX
OS Homo sapiens.
XX
FN WO2003038063-A2.
XX
PD 08-MAY-2003.
XX
PF 19-MAR-2002; 2002WO-US008277.
XX
PR 21-MAR-2001; 2001US-0277340P.
XX
PR 19-JUL-2001; 2001US-0306171P.
XX
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-430516/40.
XX
DR P-PSDB; ADC74238.
XX
PT New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).
XX
PS Claim 27; SEQ ID NO 256; 2272pp; English.
XX
CC The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein-related DNA of
CC the invention.
XX
SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
||||| ||||| ||||| ||||| |||||
Db      1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585

RESULT 43
ADC74681
ID ADC74681 standard; DNA; 2128 BP.
XX
```

```
AC ADC74681;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human secreted protein-related DNA - SEQ ID 1314.
XX
XX antianaemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
KW antidiabetic; immunosuppressive; dermatological; nephrotropic;
KW antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
KW fungicide; antiparasitic; antiarteriosclerotic; vulnerary; cytostatic;
KW haemopoietic; haematologic; anaemia; autoimmune disorder;
KW rheumatoid arthritis; inflammation; Grave's disease; diabetes;
KW systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
KW Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
KW cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
KW human; ds.
XX
XX Homo sapiens.
OS
XX WO2003038063-A2.
XX
XX 08-MAY-2003.
XX
XX 19-MAR-2002; 2002WO-US008277.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX
XX 19-JUL-2001; 2001US-0306171P.
XX
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-430516/40.
XX
XX New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).
XX
XX Disclosure; SEQ ID NO 1314; 2272pp; English.
XX
XX The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or hematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein-related DNA of
CC the invention.
XX
XX Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
SQ
```

```
Query Match 64.7%; Score 19.4; DB 10; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 1556 CTGGATCTTGGCCCGGAACCTGCATTC 1584
```

```
RESULT 44
ADD37643
ID ADD37643 standard; cDNA; 2128 BP.
```

```
XX ADD37643;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human secreted protein encoding sequence #125.
XX
XX human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
KW Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic; ss.
XX
XX Homo sapiens.
OS
XX WO200290526-A2.
XX
XX 14-NOV-2002.
XX
XX 19-MAR-2002; 2002WO-US008279.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX
XX 19-JUL-2001; 2001US-0306171P.
XX
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-140218/13.
XX
XX New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing or
PT treating allergic or asthmatic disorders, or related immediate
PT hypersensitivity disorders.
XX
XX Claim 7; SEQ ID NO 125; 1323pp; English.
XX
XX The present invention relates to an isolated polypeptide or human
CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
CC their fragments, and agonists or antagonists that bind are useful for
CC preparing a diagnostic or pharmaceutical composition for diagnosing or
CC treating allergic or asthmatic disorders. The polypeptide is also useful
CC for identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases the activity of the polypeptide. The polypeptides and nucleic
CC acid molecules are also useful for detecting, preventing, diagnosing,
CC prognosticating, treating or ameliorating inflammatory disorders
CC neoplastic diseases, wound healing and disorders of epithelial cell
CC proliferation, immune disorders, cardiovascular disorders, blood-related
CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
CC disorders. The nucleic acids are also useful for chromosome
CC identification, radiation hybrid mapping or long-range restriction
CC mapping, as molecular weight markers, or as hybridization or diagnostic
CC probes. The polypeptides and antibodies are useful for providing
CC immunological probes for differential identification of the tissues
CC immunohistochemistry assays. The present sequence represents a human
CC secreted protein encoding sequence.
XX
XX Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
SQ
```

```
Query Match 64.7%; Score 19.4; DB 10; Length 2128;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585
```

```
RESULT 45
ADD38136
ID ADD38136 standard; cDNA; 2128 BP.
```

```
XX ADD38136;
XX
```

DT 15-JAN-2004 (first entry)
 XX cDNA clone in ATCC deposit #30.
 DE human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
 XX Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic; ss.
 KW Homo sapiens.
 OS WO200290526-A2.
 PN 14-NOV-2002.
 XX 19-MAR-2002; 2002WO-US008279.
 PF 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI; 2003-140218/13.
 XX New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing or
 PT treating allergic or asthmatic disorders, or related immediate
 PT hypersensitivity disorders.
 XX Claim 1; SEQ ID NO 618; 1323pp; English.
 PS The present invention relates to an isolated polypeptide or human
 CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
 CC their fragments, and agonists or antagonists that bind are useful for
 CC preparing a diagnostic or pharmaceutical composition for diagnosing or
 CC treating allergic or asthmatic disorders. The polypeptide is also useful
 CC for identifying a binding partner by contacting the polypeptide with a
 CC binding partner, and determining whether the binding partner increases or
 CC decreases the activity of the polypeptide. The polypeptides and nucleic
 CC acid molecules are also useful for detecting, preventing, diagnosing,
 CC prognosticating, treating or ameliorating inflammatory disorders
 CC neoplastic diseases, wound healing and disorders of epithelial cell
 CC proliferation, immune disorders, cardiovascular disorders, blood-related
 CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
 CC disorders. The nucleic acids are also useful for chromosome
 CC identification, radiation hybrid mapping or long-range restriction
 CC mapping, as molecular weight markers, or as hybridization or diagnostic
 CC probes. The polypeptides and antibodies are useful for providing
 CC immunological probes for differential identification of the tissues
 CC immunohistochemistry assays. The present sequence represents a cDNA clone
 CC from ATCC deposit.
 XX SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
 Query Match 64.7%; Score 19.4; DB 10; Length 2128;
 Best Local Similarity 79.3%; Pred. No. 3.1e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
 Db 1556 CTGGATCTTGGCCCGGAACCTGCATTC 1584
 RESULT 46
 ID ADA56194
 XX ADA56194 standard; DNA; 2128 BP.
 AC ADA56194;
 XX 20-NOV-2003 (first entry)
 DT Gene encoding human secreted protein #373.
 XX DE

XX immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KW cystostatic; cerebroprotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology; ds;
 KW gene.
 XX OS Homo sapiens.
 XX WO2002102994-A2.
 XX 27-DEC-2002.
 PD 19-MAR-2002; 2002WO-US008278.
 PF 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI; 2003-167512/16.
 DR P-PSDB; ADA57090.
 XX New human secreted polypeptides and polynucleotides, useful for
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
 PT conditions, respiratory disorders, cancers, CNS disorders, or
 PT neurodegenerative disorders.
 PS Claim 21; SEQ ID NO 383; 1754pp; English.
 CC The invention relates to 592 new human secreted polypeptides useful for
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
 CC conditions, respiratory disorders, cancers, CNS disorders, or
 CC neurodegenerative disorders, or polypeptides comprising an amino acid
 CC sequence at least 93% identical to the new sequences. The polypeptides,
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic
 CC acids encoding the polypeptides, agonists or antagonists that binds to
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
 CC compositions for diagnosing, treating or preventing an e.g. immune
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease) and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
 CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to a gene encoding one of the polypeptide of the invention. Note: The
 CC sequence data for this patent did form part of the printed specification,
 CC but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;
 Query Match 64.7%; Score 19.4; DB 10; Length 2128;
 Best Local Similarity 79.3%; Pred. No. 3.1e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
 DB 1557 CTGGATCCTTGGCCCGGAACCTGCAATTC 1585

RESULT 47
 ADA57735
 ID ADA57735 standard; DNA; 2128 BP.
 XX
 AC ADA57735;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE BAC fragment containing human secreted protein gene #393.
 XX
 KW immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KW cyotatic; cerebrotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology; ds;
 KW gene; bacterial artificial chromosome.
 XX
 OS Homo sapiens.
 XX
 PN WO2002102994-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008278.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-167512/16.
 XX
 PT New human secreted polypeptides and polynucleotides, useful for
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
 PT conditions, respiratory disorders, cancers, CNS disorders, or
 PT neurodegenerative disorders.
 XX
 PS Disclosure; SEQ ID NO 1928; 1754pp; English.
 XX
 CC The invention relates to 592 new human secreted polypeptides useful for
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
 CC conditions, respiratory disorders, cancers, CNS disorders, or
 CC neurodegenerative disorders, or polypeptides comprising an amino acid
 CC sequence at least 95% identical to the new sequences. The polypeptides,
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic
 CC acids encoding the polypeptides, agonists or antagonists that binds to
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
 CC compositions for diagnosing, treating or preventing an e.g. immune
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to a bacterial artificial chromosome (BAC) fragment containing the gene
 CC encoding one of the polypeptides of the invention. Note: The sequence
 CC data for this patent did form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2128 BP; 341 A; 728 C; 603 G; 456 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 10; Length 2128;
 Best Local Similarity 79.3%; Pred. No. 3.1e-02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
 DB 1556 CTGGATCCTTGGCCCGGAACCTGCAATTC 1584

RESULT 48
 ACC50921
 ID ACC50921 standard; cDNA; 2129 BP.
 XX
 AC ACC50921;
 XX
 DT 12-JUN-2003 (first entry)
 XX
 DE Human secreted protein BAC clone SEQ ID NO 1101.
 XX
 KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
 KW vulnery; antiinflammatory; nootropic; neuroprotective;
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200295010-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 19-MAR-2002; 2002WO-US009785.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-129429/12.
 XX
 PT Novel human secreted proteins, useful for detecting, preventing,
 PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
 PT disorders such as arrhythmia.
 XX
 PS Disclosure; SEQ ID NO 1101; 1881pp; English.
 XX
 CC The present invention relates to novel human secreted proteins (ABR47633-
 CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
 CC and their coding sequences are useful for the preparation of a diagnostic
 CC or pharmaceutical composition for diagnosing or treating a cardiovascular
 CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
 CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,
 CC proliferative disorders and/or cancerous diseases and conditions, for
 CC wound healing and epithelial cell proliferation, to treat inflammation or
 CC infection, for treating thrombosis and arteriosclerosis, for treating or
 CC preventing neural damage which occurs in neuronal disorders or
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
 CC disease, to enhance bone and periodontal regeneration and aid in tissue

transplants or bone grafts, to prevent skin aging or hair loss, to stimulate growth and differentiation of haematopoietic cells and bone marrow cells when used in combination with other cytokines, to maintain organs before transplantation or for supporting cell culture of primary tissues, to increase or decrease differentiation or proliferation of embryonic stem cells, or to modulate mammalian characteristics or metabolism. The present sequence was used to illustrate the invention. Note: The sequence data for this patent was published in electronic format and is available from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2129 BP; 354 A; 716 C; 597 G; 462 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 8; Length 2129;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
||||| ||||||| |||||||
Db 1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585

RESULT 49
ABZ71523
ID ABZ71523 standard; DNA; 2129 BP.
XX AC ABZ71523;
XX DT 03-APR-2003 (first entry)
XX DE Secreted protein gene 118 genomic fragment HNGO112, SEQ ID NO:633.
XX KW Human; secreted protein; digestive disorder; gastrointestinal disorder;
KW mouth; oesophagus; stomach; small intestine; large intestine; liver;
KW biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
KW immune disorder; inflammation; infection; wound healing; drug screening;
KW chromosome identification; chromosome mapping; cytostatic; gene therapy;
KW antiinflammatory; immunosuppressive; vulnery; chromosome 11; gene; da.
XX OS Homo sapiens.
XX PN WO200276488-A1.
XX PD 03-OCT-2002.
XX PF 19-MAR-2002; 2002WO-US008276.
XX XX 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;
XX PT WPI; 2003-029900/02.
XX DR
XX PT New human secreted proteins and nucleic acids, useful for detecting,
PT preventing, diagnosing, prognosticating, treating and/or ameliorating
PT e.g. gastrointestinal diseases and disorders, or cancers.
XX PS Disclosure; Page 1190-1191; 1216pp; English.

XX CC ABZ71190-ABZ71478 represent cDNAs corresponding to 178 human secreted
CC protein genes, and ABP00011-ABP00299 represent the proteins they encode.
CC CC ABZ71479-ABZ71540 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening, and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing, treating, ameliorating or preventing
CC digestive disorders. Such conditions include disorders of the mouth,

CC oesophagus, stomach, small intestine, large intestine, liver, biliary
CC tract and pancreas, and include cancers of these organs and tissues. The
CC secreted proteins and their nucleic acids may also be used in the
CC treatment of immune disorders, inflammation, infection,
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids
CC of the invention may be used for chromosome identification, chromosome
CC mapping, in gene therapy, for identifying individuals from minute
CC biological samples, as hybridisation probes, and as molecular weight
CC markers. The present sequence represents a human secreted protein genomic
CC fragment referred to in the disclosure of the invention

XX SQ Sequence 2129 BP; 354 A; 716 C; 597 G; 462 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 8; Length 2129;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 29
||||| ||||||| |||||||
Db 1557 CTGGATCTTGGCCCGGAACCTGCATTC 1585

RESULT 50
ADB91884
ID ADB91884 standard; DNA; 2129 BP.
XX AC ADB91884;
XX DT 04-DEC-2003 (first entry)
XX DE Human secreted protein related DNA #SEQ ID 830.
XX KW Secreted protein; gene therapy; antidiabetic; diabetes; human; da.
XX OS Homo sapiens.
XX PN WO2003004622-A2.
XX PD 16-JAN-2003.
XX PF 19-MAR-2002; 2002WO-US008124.
XX PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX PT WPI; 2003-229407/22.

XX PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
XX PS Disclosure; SEQ ID NO 830; 1537pp; English.

XX CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2129 BP; 354 A; 716 C; 597 G; 462 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 9; Length 2129;
Best Local Similarity 79.3%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 CTGGATCCTTGCCCGCCCGCTTGAAATCC 29
Db 1557 CTGGATTCTTGCCCGCGGAACCTGCATTCC 1585

Search completed: April 25, 2005, 11:36:00
Job time : 324.224 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:13:24 ; Search time 2581.55 Seconds
(without alignments)
442.342 Million cell updates/sec

Title: US-10-010-476-10
Perfect score: 30
Sequence: 1 CTGGATCTTGCCCGCCCTTGAATTCCTCC 30

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database: EST:

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	76.7	1240	8	CC249763 CH261-16E
2	22	73.3	691	2	BB643723
3	21.6	72.0	601	7	CK840615 UI-R-AFI-
4	21.6	72.0	695	9	CE112892 tigr-g88-
5	21.6	72.0	1164	5	BQ953457 AGENCOURT
6	21.2	70.7	207	2	BB593516
7	21	70.0	507	2	BB753706 BB753706
8	21	70.0	540	5	BY412408 BY412408
9	21	70.0	757	7	CK773708 962382 MA
10	21	70.0	1139	1	AL567601 AL567601
11	20.8	69.3	1158	6	CD497293 CDA27-A09
12	20.6	68.7	565	7	CN704387 E0487A12
13	20.6	68.7	654	7	CN698328
14	20.6	68.7	3697	3	AK039759 Mus muscu
15	20.4	68.0	378	1	AA763854 v43h01.r
16	20.4	68.0	386	6	BY605779 BY605779
17	20.4	68.0	394	1	AI428664 v43h01.y
18	20.4	68.0	499	1	AA320450 v22609.r
19	20.4	68.0	529	8	AZ285328
20	20.4	68.0	544	2	BB730755 BB730755
21	20.4	68.0	594	7	CO881510 BovGen_09
22	20.4	68.0	691	9	CE337240 tigr-g88-
23	20.4	68.0	724	4	BI555868 603218173
24	20.4	68.0	725	3	BC028653 Mus muscu

25	20.4	68.0	733	9	AG441314
26	20.4	68.0	747	9	AG413802 Mus muscu
27	20.4	68.0	753	9	CR201952 Forward s
28	20.4	68.0	803	9	CR124383 Forward s
29	20.4	68.0	1039	9	AG082767 Pan trogl
30	20	66.7	340	1	AI646086 v45402.x
31	20	66.7	373	7	W61990 md83b02.r1
32	20	66.7	417	7	W75418 me50b02.r1
33	20	66.7	511	2	BB283460 BB283460
34	20	66.7	517	4	BG067160 H3051C07-
35	20	66.7	520	6	BA884170 B0108D11-
36	20	66.7	528	5	BQ523276 NISC nl21
37	20	66.7	560	1	AI225464 us87h11.y
38	20	66.7	576	2	BE289377
39	20	66.7	661	8	AZ744176
40	20	66.7	674	6	BY757487 BY757487
41	20	66.7	680	9	CR085351 Forward s
42	20	66.7	684	2	AW187871 BNLGH1134
43	20	66.7	688	9	CR047114 Forward s
44	20	66.7	889	9	CG436661 OGTCG93TV
45	20	66.7	890	4	BI525356 60294465
46	20	66.7	890	9	CL126864 ISB1-9111
47	20	66.7	1015	7	CP223569 AGENCOURT
48	20	66.7	1537	8	CC237106 CH261-191
49	19.8	65.3	669	5	BX308381 BX308381
50	19.8	65.3	714	6	CB509377 ssalwH50
51	19.6	65.3	283	7	CF094505 QHN18C01.
52	19.6	65.3	389	7	CV498081 62335.1 M
53	19.6	65.3	422	1	AU042204 AU042204
54	19.6	65.3	469	8	BZ816795 PUGAV67TD
55	19.6	65.3	481	6	CA516654 KS09061FB
56	19.6	65.3	489	8	BZ816787 PUGAV67TB
57	19.6	65.3	511	5	BQ142790 fml1C.pk0
58	19.6	65.3	513	2	AW096805 EST289985
59	19.6	65.3	543	5	BQ120494 EST606070
60	19.6	65.3	555	4	BG589881 EST497723
61	19.6	65.3	556	4	BG128705 EST474351
62	19.6	65.3	570	2	AW093131 EST286311
63	19.6	65.3	590	4	BG072305 H3109A10-
64	19.6	65.3	590	7	CV501298 65800.1 M
65	19.6	65.3	599	5	BQ120493 EST606069
66	19.6	65.3	604	4	BG085139 H3109A10-
67	19.6	65.3	605	4	BI421317 EST531983
68	19.6	65.3	607	4	BI924690 EST544579
69	19.6	65.3	618	5	BQ045585 EST594703
70	19.6	65.3	622	7	CV500084 64507.1 M
71	19.6	65.3	624	1	AI777259 EST258224
72	19.6	65.3	638	1	AJ804621 AJ804621
73	19.6	65.3	649	8	BZ020499 oeh06e01.
74	19.6	65.3	666	4	BI924873 EST544762
75	19.6	65.3	696	4	BG125827 EST471473
76	19.6	65.3	705	4	BG643745 EST511939
77	19.6	65.3	792	7	CK483353 AGENCOURT
78	19.6	65.3	797	7	CV499373 63722.1 M
79	19.6	65.3	816	2	BF676759 602086433
80	19.6	65.3	819	2	BF262066 HV CEA000
81	19.4	64.7	119	6	CB106260 K-EST0138
82	19.4	64.7	172	8	AQ346229 RPC111-11
83	19.4	64.7	269	4	BM208321 C0625F08-
84	19.4	64.7	271	2	BB608158 BB608158
85	19.4	64.7	326	7	T38495 EST103992 S
86	19.4	64.7	331	5	BM222541 K0122G08-
87	19.4	64.7	337	4	BY393360 BY393360
88	19.4	64.7	348	6	CB704661 AMGNNUC.N
89	19.4	64.7	352	1	AA032751 m135a02.r
90	19.4	64.7	357	5	BY400112 BY400112
91	19.4	64.7	360	9	CE833718 tigr-g88-
92	19.4	64.7	361	5	BY467320 BY467320
93	19.4	64.7	366	5	BY411617 BY411617
94	19.4	64.7	370	5	BY405025 BY405025
95	19.4	64.7	372	6	CD315499 StrRu621.
96	19.4	64.7	376	6	BY681852 BY681852
97	19.4	64.7	379	1	AA260010 va87g06.r

98	19.4	64.7	380	5	BY366757	BY366757	171	19.4	64.7	696	6	BY762798	BY762798
99	19.4	64.7	383	2	BB837024	BB837024	c 172	19.4	64.7	696	6	CD835459	CD835459
c 100	19.4	64.7	383	4	BM208863	BM208863	173	19.4	64.7	696	8	BH021054	BH021054
101	19.4	64.7	384	1	AA763480	AA763480	174	19.4	64.7	716	6	CB204291	CB204291
102	19.4	64.7	385	6	BY677468	BY677468	c 175	19.4	64.7	733	5	BX862503	BX862503
103	19.4	64.7	387	1	AI036434	AI036434	c 176	19.4	64.7	739	4	BG064967	BG064967
104	19.4	64.7	387	5	BY409902	BY409902	c 177	19.4	64.7	746	1	AA986905	AA986905
105	19.4	64.7	387	6	BY701593	BY701593	c 178	19.4	64.7	769	7	CN248508	CN248508
106	19.4	64.7	392	5	BY440840	BY440840	c 179	19.4	64.7	772	6	CD826862	CD826862
107	19.4	64.7	396	2	BB794167	BB794167	c 180	19.4	64.7	772	9	AG517523	AG517523
108	19.4	64.7	401	7	W88143	W88143	c 181	19.4	64.7	786	5	BX314896	BX314896
c 109	19.4	64.7	412	4	BM119162	BM119162	c 182	19.4	64.7	788	6	CA383229	CA383229
c 110	19.4	64.7	413	4	BM224038	BM224038	183	19.4	64.7	805	5	BU237407	BU237407
c 111	19.4	64.7	418	4	BM224072	BM224072	184	19.4	64.7	807	2	BF685564	BF685564
c 112	19.4	64.7	420	5	BY392061	BY392061	185	19.4	64.7	904	5	BO216598	BO216598
c 113	19.4	64.7	422	4	BM195973	BM195973	c 186	19.4	64.7	921	7	CF582656	CF582656
c 114	19.4	64.7	422	4	BM220960	BM220960	187	19.4	64.7	953	7	CF582656	CF582656
c 115	19.4	64.7	429	2	BM731596	BM731596	188	19.4	64.7	1157	9	CL051085	CL051085
c 116	19.4	64.7	437	7	W89731	W89731	189	19.4	64.7	1183	9	CL079616	CL079616
c 117	19.4	64.7	452	5	BY474650	BY474650	190	19.4	64.7	1192	9	AG350284	AG350284
c 118	19.4	64.7	455	2	BB728089	BB728089	191	19.4	64.7	1277	9	CL023471	CL023471
c 119	19.4	64.7	459	5	BY484207	BY484207	192	19.4	64.7	1314	9	AG347170	AG347170
c 120	19.4	64.7	460	5	BQ745034	BQ745034	c 193	19.2	64.0	337	1	AA228328	AA228328
c 121	19.4	64.7	460	5	BY472485	BY472485	194	19.2	64.0	381	7	CO093571	CO093571
c 122	19.4	64.7	462	5	BY472475	BY472475	195	19.2	64.0	423	8	AZ484723	AZ484723
c 123	19.4	64.7	470	1	AI153827	AI153827	c 196	19.2	64.0	437	5	BY006603	BY006603
c 124	19.4	64.7	472	4	BM235089	BM235089	197	19.2	64.0	457	7	N98614	N98614
c 125	19.4	64.7	476	8	BZ837656	BZ837656	c 198	19.2	64.0	461	9	CL356120	CL356120
c 126	19.4	64.7	477	8	AQ668168	AQ668168	199	19.2	64.0	471	1	AJ668382	AJ668382
c 127	19.4	64.7	480	2	BB771345	BB771345	c 200	19.2	64.0	493	9	CE565055	CE565055
c 128	19.4	64.7	487	1	AI847905	AI847905	c 201	19.2	64.0	584	8	AQ609416	AQ609416
c 129	19.4	64.7	487	6	CA894579	CA894579	c 202	19.2	64.0	603	4	BJ195514	BJ195514
c 130	19.4	64.7	491	4	BM221577	BM221577	203	19.2	64.0	606	4	BJ604059	BJ604059
c 131	19.4	64.7	492	1	AA014622	AA014622	204	19.2	64.0	619	9	BX960595	BX960595
c 132	19.4	64.7	494	5	BY467013	BY467013	205	19.2	64.0	627	4	BY590667	BY590667
c 133	19.4	64.7	496	5	BY465974	BY465974	206	19.2	64.0	630	4	BJ172664	BJ172664
c 134	19.4	64.7	506	5	BY470273	BY470273	207	19.2	64.0	637	4	BJ590311	BJ590311
c 135	19.4	64.7	507	4	BM220720	BM220720	208	19.2	64.0	638	4	BJ589567	BJ589567
c 136	19.4	64.7	511	2	BM757154	BM757154	209	19.2	64.0	673	4	BJ580428	BJ580428
c 137	19.4	64.7	511	4	BM237369	BM237369	210	19.2	64.0	685	4	BJ581569	BJ581569
c 138	19.4	64.7	512	2	BB761057	BB761057	211	19.2	64.0	701	4	BJ579010	BJ579010
c 139	19.4	64.7	514	5	BY471447	BY471447	212	19.2	64.0	702	4	BJ593096	BJ593096
c 140	19.4	64.7	517	2	BB755538	BB755538	213	19.2	64.0	711	4	BJ581862	BJ581862
c 141	19.4	64.7	517	2	BB768718	BB768718	214	19.2	64.0	725	4	BJ579893	BJ579893
c 142	19.4	64.7	517	5	BY480323	BY480323	215	19.2	64.0	727	4	BJ587322	BJ587322
c 143	19.4	64.7	519	5	BY470899	BY470899	216	19.2	64.0	731	4	BJ581518	BJ581518
c 144	19.4	64.7	521	7	CF198056	CF198056	217	19.2	64.0	738	4	BJ590545	BJ590545
c 145	19.4	64.7	521	7	CF198056	CF198056	218	19.2	64.0	739	4	BJ586364	BJ586364
c 146	19.4	64.7	525	5	BY471840	BY471840	219	19.2	64.0	770	4	BJ579007	BJ579007
c 147	19.4	64.7	527	5	BY474745	BY474745	220	19.2	64.0	774	9	CG917970	CG917970
c 148	19.4	64.7	532	4	BM247664	BM247664	221	19.2	64.0	775	4	BJ583282	BJ583282
c 149	19.4	64.7	542	5	BY479275	BY479275	222	19.2	64.0	775	4	BJ605785	BJ605785
c 150	19.4	64.7	546	6	BM222271	BM222271	223	19.2	64.0	780	4	BJ594593	BJ594593
c 151	19.4	64.7	556	6	CA892769	CA892769	224	19.2	64.0	784	4	BJ590265	BJ590265
c 152	19.4	64.7	558	1	AA500393	AA500393	225	19.2	64.0	798	4	BJ583368	BJ583368
c 153	19.4	64.7	561	1	AA500393	AA500393	c 226	19.2	64.0	804	9	AG532109	AG532109
c 154	19.4	64.7	565	4	BM119657	BM119657	227	19.2	64.0	844	9	CNS04K07	CNS04K07
c 155	19.4	64.7	571	4	BM223834	BM223834	228	19.2	64.0	848	7	CF375823	CF375823
c 156	19.4	64.7	572	4	BM205153	BM205153	229	19.2	64.0	992	9	CNS050P0	CNS050P0
c 157	19.4	64.7	587	4	BI646002	BI646002	230	19.2	64.0	1003	5	BQ214771	BQ214771
c 158	19.4	64.7	587	4	BM250087	BM250087	231	19.2	64.0	1058	9	CNS05NNG	CNS05NNG
c 159	19.4	64.7	613	7	CK347000	CK347000	232	19.2	64.0	1088	5	BQ944517	BQ944517
c 160	19.4	64.7	619	5	BP767583	BP767583	233	19.2	64.0	1101	9	CNS001TV	CNS001TV
c 161	19.4	64.7	620	6	CA318800	CA318800	c 234	19.2	64.0	1101	9	CNS00H4U	CNS00H4U
c 162	19.4	64.7	625	6	CB522815	CB522815	235	19	63.3	279	9	AG592583	AG592583
c 163	19.4	64.7	629	2	AW536490	AW536490	c 236	19	63.3	329	9	CG803303	CG803303
c 164	19.4	64.7	638	7	CK764131	CK764131	237	19	63.3	345	9	CG733119	CG733119
c 165	19.4	64.7	644	5	BO747303	BO747303	238	19	63.3	347	8	AQ644387	AQ644387
c 166	19.4	64.7	655	7	BX314895	BX314895	239	19	63.3	405	1	AU301652	AU301652
c 167	19.4	64.7	657	7	CF902322	CF902322	c 240	19	63.3	408	8	AQ644025	AQ644025
c 168	19.4	64.7	663	5	BX314046	BX314046	241	19	63.3	450	7	T92440	T92440
c 169	19.4	64.7	679	9	AG311683	AG311683	c 242	19	63.3	477	1	AU045564	AU045564
c 170	19.4	64.7	680	7	CO043288	CO043288	c 243	19	63.3	482	5	BY314544	BY314544

244	19	63.3	503	5	BY476139	BY476139	BY476139	317	18.8	62.7	420	5	BY639531	BY639531
C 245	19	63.3	508	7	CK885684	CK885684	CK885684	C 318	18.8	62.7	422	6	BY624157	BY624157
246	19	63.3	528	9	CG597427	CG597427	CG597427	C 319	18.8	62.7	422	6	BY655008	BY655008
247	19	63.3	535	8	TA1114D06Q	TA1114D06Q	TA1114D06Q	C 320	18.8	62.7	423	5	BY444320	BY444320
248	19	63.3	539	8	AQ941937	AQ941937	AQ941937	C 321	18.8	62.7	429	1	AI256606	AI256606
C 249	19	63.3	541	8	AQ953350	Sheared D	Sheared D	C 322	18.8	62.7	431	2	BP723493	BP723493
C 250	19	63.3	541	6	AQ953350	Sheared D	Sheared D	C 323	18.8	62.7	431	5	BP723493	BP723493
C 251	19	63.3	571	6	CB512781	CB512781	CB512781	C 324	18.8	62.7	431	5	BY421949	BY421949
C 252	19	63.3	580	4	CG800719	CG800719	CG800719	C 324	18.8	62.7	442	4	BG370491	BG370491
C 253	19	63.3	595	6	CA731113	CA731113	CA731113	C 325	18.8	62.7	450	2	BE135471	BE135471
C 254	19	63.3	639	8	AZ420700	AZ420700	AZ420700	C 326	18.8	62.7	451	2	BE741783	BE741783
C 255	19	63.3	668	8	AZ104538	AZ104538	AZ104538	C 327	18.8	62.7	456	7	R21690	R21690
C 256	19	63.3	689	9	CE819120	CE819120	CE819120	C 328	18.8	62.7	457	7	H20077	H20077
C 257	19	63.3	702	8	BP975378	BP975378	BP975378	C 329	18.8	62.7	465	2	BB760777	BB760777
C 258	19	63.3	728	9	CW016950	CW016950	CW016950	C 330	18.8	62.7	465	2	CA893983	CA893983
C 259	19	63.3	776	2	BF214319	BF214319	BF214319	C 331	18.8	62.7	468	7	CA047464	CA047464
C 260	19	63.3	790	8	BF2150261	BF2150261	BF2150261	C 332	18.8	62.7	473	6	CA531889	CA531889
C 261	19	63.3	805	7	CK196621	CK196621	CK196621	C 333	18.8	62.7	475	7	W83938	W83938
C 262	19	63.3	808	5	BO143964	BO143964	BO143964	C 334	18.8	62.7	489	9	BX992850	BX992850
C 263	19	63.3	851	5	BUE25158	BUE25158	BUE25158	C 335	18.8	62.7	497	6	CB664412	CB664412
C 264	19	63.3	858	7	CF619338	CF619338	CF619338	C 336	18.8	62.7	506	7	CNS38690	CNS38690
C 265	19	63.3	911	5	BO714137	BO714137	BO714137	C 337	18.8	62.7	514	2	BE523647	BE523647
C 266	19	63.3	933	5	BQ943145	BQ943145	BQ943145	C 338	18.8	62.7	520	5	BE523647	BE523647
C 267	19	63.3	956	4	BG536382	BG536382	BG536382	C 339	18.8	62.7	524	2	BE136308	BE136308
C 268	19	63.3	1003	9	CNS0258A	CNS0258A	CNS0258A	C 340	18.8	62.7	526	7	CR442462	CR442462
C 269	19	63.3	1008	8	CC193033	CC193033	CC193033	C 341	18.8	62.7	529	1	AA255095	AA255095
C 270	19	63.3	1111	4	BM549001	BM549001	BM549001	C 342	18.8	62.7	535	6	C79367	C79367
C 271	19	63.3	1241	8	CC293456	CC293456	CC293456	C 343	18.8	62.7	546	6	CB524103	CB524103
C 272	19	63.3	1269	4	BM474632	BM474632	BM474632	C 344	18.8	62.7	549	5	BU700349	BU700349
C 273	19	63.3	2700	3	AK029854	AK029854	AK029854	C 345	18.8	62.7	552	7	CK990730	CK990730
C 274	18.8	62.7	160	2	BB600980	BB600980	BB600980	C 346	18.8	62.7	569	2	BE870176	BE870176
C 275	18.8	62.7	196	1	AV143367	AV143367	AV143367	C 347	18.8	62.7	577	4	BG066455	BG066455
C 276	18.8	62.7	272	4	AW324515	AW324515	AW324515	C 348	18.8	62.7	580	5	BU700791	BU700791
C 277	18.8	62.7	299	2	AW324515	AW324515	AW324515	C 349	18.8	62.7	585	4	BP208448	BP208448
C 278	18.8	62.7	309	9	CE491137	CE491137	CE491137	C 350	18.8	62.7	587	5	BJ314056	BJ314056
C 279	18.8	62.7	309	9	CE491137	CE491137	CE491137	C 351	18.8	62.7	587	6	C79224	C79224
C 280	18.8	62.7	314	7	CF076555	CF076555	CF076555	C 352	18.8	62.7	597	6	CB447945	CB447945
C 281	18.8	62.7	318	1	AV212922	AV212922	AV212922	C 353	18.8	62.7	606	6	CA323006	CA323006
C 282	18.8	62.7	318	1	AW519623	AW519623	AW519623	C 354	18.8	62.7	617	6	CA370447	CA370447
C 283	18.8	62.7	332	1	AI851207	AI851207	AI851207	C 355	18.8	62.7	617	7	CNS036352	CNS036352
C 284	18.8	62.7	340	2	BF564390	BF564390	BF564390	C 356	18.8	62.7	629	7	CNS037612	CNS037612
C 285	18.8	62.7	344	2	BF720236	BF720236	BF720236	C 357	18.8	62.7	630	8	AZ438685	AZ438685
C 286	18.8	62.7	348	2	BF763546	BF763546	BF763546	C 358	18.8	62.7	635	5	BQ392943	BQ392943
C 287	18.8	62.7	348	5	BY409848	BY409848	BY409848	C 359	18.8	62.7	649	6	BY719500	BY719500
C 288	18.8	62.7	360	5	BY427355	BY427355	BY427355	C 360	18.8	62.7	656	1	AU133700	AU133700
C 289	18.8	62.7	362	5	BY427355	BY427355	BY427355	C 361	18.8	62.7	659	9	CE000883	CE000883
C 290	18.8	62.7	363	2	BB791989	BB791989	BB791989	C 362	18.8	62.7	667	4	BJ263244	BJ263244
C 291	18.8	62.7	365	6	BY664116	BY664116	BY664116	C 363	18.8	62.7	672	9	AG079418	AG079418
C 292	18.8	62.7	369	2	BB791713	BB791713	BB791713	C 364	18.8	62.7	679	2	BB036238	BB036238
C 293	18.8	62.7	374	5	BY440242	BY440242	BY440242	C 365	18.8	62.7	683	6	CB664452	CB664452
C 294	18.8	62.7	376	1	AI180640	AI180640	AI180640	C 366	18.8	62.7	684	6	CA366933	CA366933
C 295	18.8	62.7	380	1	AU040726	AU040726	AU040726	C 367	18.8	62.7	688	6	BY760890	BY760890
C 296	18.8	62.7	381	2	AW534146	AW534146	AW534146	C 368	18.8	62.7	695	6	CB450850	CB450850
C 297	18.8	62.7	385	5	BY425208	BY425208	BY425208	C 369	18.8	62.7	695	6	CB450850	CB450850
C 298	18.8	62.7	386	6	BY760678	BY760678	BY760678	C 370	18.8	62.7	698	6	CB469341	CB469341
C 299	18.8	62.7	388	6	BY661752	BY661752	BY661752	C 371	18.8	62.7	709	6	CB655224	CB655224
C 300	18.8	62.7	393	6	BY606920	BY606920	BY606920	C 372	18.8	62.7	712	6	CB655217	CB655217
C 301	18.8	62.7	394	6	BY667126	BY667126	BY667126	C 373	18.8	62.7	713	9	AG535518	AG535518
C 302	18.8	62.7	396	9	CE273353	CE273353	CE273353	C 374	18.8	62.7	717	7	CF532987	CF532987
C 303	18.8	62.7	398	4	BM345384	BM345384	BM345384	C 375	18.8	62.7	725	7	CO813325	CO813325
C 304	18.8	62.7	401	1	AI060910	AI060910	AI060910	C 376	18.8	62.7	729	2	BF037401	BF037401
C 305	18.8	62.7	404	9	CNS02770B	CNS02770B	CNS02770B	C 377	18.8	62.7	730	8	AZ997891	AZ997891
C 306	18.8	62.7	404	5	BY427799	BY427799	BY427799	C 378	18.8	62.7	731	9	AG535518	AG535518
C 307	18.8	62.7	405	2	BB788890	BB788890	BB788890	C 379	18.8	62.7	733	9	CG127526	CG127526
C 308	18.8	62.7	407	1	AI528582	AI528582	AI528582	C 380	18.8	62.7	739	6	CB669408	CB669408
C 309	18.8	62.7	407	2	BB683981	BB683981	BB683981	C 381	18.8	62.7	745	7	CO115005	CO115005
C 310	18.8	62.7	411	7	R81401	R81401	R81401	C 382	18.8	62.7	755	9	AG418108	AG418108
C 311	18.8	62.7	413	4	BI961920	BI961920	BI961920	C 383	18.8	62.7	759	4	BI913074	BI913074
C 312	18.8	62.7	414	6	BF661462	BF661462	BF661462	C 384	18.8	62.7	764	4	BG923836	BG923836
C 313	18.8	62.7	416	2	BF463794	BF463794	BF463794	C 385	18.8	62.7	772	6	CB674490	CB674490
C 314	18.8	62.7	417	8	AQ024128	AQ024128	AQ024128	C 386	18.8	62.7	788	5	BX297790	BX297790
C 315	18.8	62.7	418	1	AI266350	AI266350	AI266350	C 387	18.8	62.7	802	6	CB651772	CB651772
C 316	18.8	62.7	418	2	BB686634	BB686634	BB686634	C 388	18.8	62.7	803	7	CF131592	CF131592
								C 389	18.8	62.7	820	7	CK020737	CK020737

C 390	18.8	62.7	826	4	BG778643	602667908	463	18.6	62.0	514	7	C0722905	C0722905	Mdbb5001a
C 391	18.8	62.7	826	2	BF674866	602136733	C 464	18.6	62.0	520	4	BM258031	52258031	MA
C 392	18.8	62.7	837	6	CD793241	EST664602	C 465	18.6	62.0	522	6	CA427653	UI-H-DF0-	UI-H-DF0-
C 393	18.8	62.7	849	4	BM047088	603827419	C 466	18.6	62.0	529	2	BE755504	209361	MA
C 394	18.8	62.7	849	4	CC913661	t079m18da	C 467	18.6	62.0	535	5	BQ447070	UI-H-EU1-	UI-H-EU1-
C 395	18.8	62.7	885	8	BH160563	ENT0W73TR	C 468	18.6	62.0	535	6	CA418441	UI-H-EU1-	UI-H-EU1-
C 396	18.8	62.7	895	5	BX329310	EX329310	C 469	18.6	62.0	536	5	BQ446891	UI-H-EU1-	UI-H-EU1-
C 397	18.8	62.7	899	6	CA978594	AGENCOURT	C 470	18.6	62.0	538	2	AW107335	um14e11.x	AW107335
C 398	18.8	62.7	904	6	CB196315	AGENCOURT	C 471	18.6	62.0	569	1	AV615651	AV615651	AV615651
C 399	18.8	62.7	905	5	BU543512	AGENCOURT	C 472	18.6	62.0	575	4	BG799197	fp31e07.y	BG799197
C 400	18.8	62.7	911	7	CF618648	AGENCOURT	C 473	18.6	62.0	575	6	CB442838	693667	MA
C 401	18.8	62.7	916	5	BU588797	AGENCOURT	C 474	18.6	62.0	578	6	CB422974	596139	MA
C 402	18.8	62.7	919	6	CA977913	AGENCOURT	C 475	18.6	62.0	607	6	CB440801	691108	MA
C 403	18.8	62.7	927	5	BUI146217	AGENCOURT	C 476	18.6	62.0	612	5	BQ573868	UI-H-EU1-	UI-H-EU1-
C 404	18.8	62.7	961	4	BI521504	603081321	C 477	18.6	62.0	615	5	BQ447130	UI-H-EU1-	UI-H-EU1-
C 405	18.8	62.7	971	8	CC252450	CH261-150	C 478	18.6	62.0	628	9	CE635517	tigr-g88-	tigr-g88-
C 406	18.8	62.7	982	9	CL146314	ISB1-147A	C 479	18.6	62.0	644	8	AZ390219	1M0151P22	AZ390219
C 407	18.8	62.7	985	9	AG363871	Mus muscu	C 480	18.6	62.0	653	5	BQ006783	UI-H-EU1-	UI-H-EU1-
C 408	18.8	62.7	1032	5	BQ069706	AGENCOURT	C 481	18.6	62.0	673	5	BQ445900	UI-H-EU1-	UI-H-EU1-
C 409	18.8	62.7	1036	5	BQ067958	AGENCOURT	C 482	18.6	62.0	653	7	CK832072	4055728	B
C 410	18.8	62.7	1064	9	CL017404	CH216-2B2	C 483	18.6	62.0	661	5	BQ182753	UI-H-EU0-	UI-H-EU0-
C 411	18.8	62.7	1116	4	BM563927	AGENCOURT	C 484	18.6	62.0	665	5	BQ616718	UI-H-DF0-	UI-H-DF0-
C 412	18.8	62.7	1146	9	CL041995	CH216-53L	C 485	18.6	62.0	667	7	CN788956	4123348	B
C 413	18.8	62.7	1189	9	CL081402	CH216-161	C 486	18.6	62.0	673	4	BG926371	HNC55-1-H	BG926371
C 414	18.8	62.7	1192	4	BGI22021	602351683	C 487	18.6	62.0	685	5	BQ044947	UI-H-EU0-	UI-H-EU0-
C 415	18.8	62.7	1200	4	BM471735	AGENCOURT	C 488	18.6	62.0	690	8	BZ509608	BOMOL61TF	BZ509608
C 416	18.8	62.7	1201	4	BM545837	AGENCOURT	C 489	18.6	62.0	695	5	BUI61709	UI-H-DF0-	UI-H-DF0-
C 417	18.8	62.7	1255	9	CL649095	CH213-214	C 490	18.6	62.0	703	5	BQ183342	UI-H-EU0-	UI-H-EU0-
C 418	18.8	62.7	1417	5	BQ895239	AGENCOURT	C 491	18.6	62.0	703	5	BQ771633	UI-H-EU1-	UI-H-EU1-
C 419	18.8	62.7	1465	3	AK011553	Mus muscu	C 492	18.6	62.0	703	8	BH959217	odj10d01.	BH959217
C 420	18.8	62.7	1656	3	AK077841	Mus muscu	C 493	18.6	62.0	710	5	BQ183771	UI-H-EU0-	UI-H-EU0-
C 421	18.8	62.0	205	9	CE653978	tigr-g88-	C 494	18.6	62.0	728	8	BZ054695	1nr33d09.	BZ054695
C 422	18.6	62.0	253	1	AI470540	ti90g11.x	C 495	18.6	62.0	728	9	AG596292	Mus muscu	AG596292
C 423	18.6	62.0	263	8	BH496852	BOHLA25TF	C 496	18.6	62.0	748	9	AG5229526	Mus muscu	AG5229526
C 424	18.6	62.0	275	5	BQ006680	UI-H-EU1-	C 497	18.6	62.0	750	7	C0875776	Bovden_04	C0875776
C 425	18.6	62.0	275	9	CE181828	tigr-g88-	C 498	18.6	62.0	762	9	AGS27238	Mus muscu	AGS27238
C 426	18.6	62.0	280	6	CA412761	UI-H-EU0-	C 499	18.6	62.0	763	8	BH699927	BOMNZ41TF	BH699927
C 427	18.6	62.0	281	4	BM431733	1duo26F11	C 500	18.6	62.0	769	7	CK448232	pnc9c915aP	CK448232
C 428	18.6	62.0	283	5	BQ447114	UI-H-EU1-								
C 429	18.6	62.0	289	6	CA419191	UI-H-EU1-								
C 430	18.6	62.0	294	5	BQ181527	UI-H-EU0-								
C 431	18.6	62.0	338	1	AA959728	vw54c07.8								
C 432	18.6	62.0	339	1	AA959831	vw54g07.8								
C 433	18.6	62.0	339	6	CD309742	StrPu691.								
C 434	18.6	62.0	344	7	CN480369	UI-H-EU0-								
C 435	18.6	62.0	345	5	BM988025	UI-H-DF0-								
C 436	18.6	62.0	361	9	CC791686	ZMMEBb016								
C 437	18.6	62.0	384	5	CO599170	DG8-177p1								
C 438	18.6	62.0	386	5	BQ447790	UI-H-EU1-								
C 439	18.6	62.0	390	2	AW207201	UI-H-BI1-								
C 440	18.6	62.0	398	2	BE218409	hV38c07.x								
C 441	18.6	62.0	402	2	BF196113	7n46d01.x								
C 442	18.6	62.0	411	1	AI684377	tc96f11.x								
C 443	18.6	62.0	413	1	BQ029011	UI-H-DF0-								
C 444	18.6	62.0	416	5	BQ574629	UI-H-EU1-								
C 445	18.6	62.0	442	5	BV366980	BV366980								
C 446	18.6	62.0	446	5	BF112731	BP112731								
C 447	18.6	62.0	459	2	BF192820	244178								
C 448	18.6	62.0	466	7	MA33246	SWJ3CA682SK								
C 449	18.6	62.0	467	1	AA937215	ok13f05.8								
C 450	18.6	62.0	470	5	BQ447743	UI-H-EU1-								
C 451	18.6	62.0	471	5	BQ447819	UI-H-EU1-								
C 452	18.6	62.0	471	5	BUI617886	UI-H-DF0-								
C 453	18.6	62.0	472	5	BUI617986	UI-H-DF0-								
C 454	18.6	62.0	477	5	BQ182663	UI-H-EU0-								
C 455	18.6	62.0	487	5	BP112595	BP112595								
C 456	18.6	62.0	490	6	CA413462	UI-H-EU0-								
C 457	18.6	62.0	491	2	BE481500	166971								
C 458	18.6	62.0	491	6	CA428195	UI-H-DF0-								
C 459	18.6	62.0	493	6	CA427568	UI-H-DF0-								
C 460	18.6	62.0	494	5	BX095415	BX095415								
C 461	18.6	62.0	507	5	BUI617692	UI-H-DF0-								
C 462	18.6	62.0	512	6	CA668721	w1eul.p40								

ALIGNMENTS

1240 bp DNA linear GSS 13-MAY-2003

CH261-16B9_Sp6.1 CH261 Gallus gallus genomic clone CH261-16B9,

genomic survey sequence.

CC249763

CC249763.1 GI:30586513

GSS

Gallus gallus (chicken)

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus

1 (bases 1 to 1240)

Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,

Warren,W., Graves,T., Mardis,E. and Wilson,R.

Gallus gallus BAC End Reads

Unpublished (2003)

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Insert Length: 182000 Std Error: 0.00

Seq primer: Sp6 ATTTAGGTGACACTATAG

Class: BAC ends

High quality sequence start: 20

High quality sequence stop: 627.

Location/Qualifiers

1. .1240

/organism="Gallus gallus"

RESULT 1

CC249763

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source


```

/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-1659"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match 76.7%; Score 23; DB 8; Length 1240;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGATCTTGCCCGCCCGCTTG 23
|||||
Db 869 CTGGATCTTGCCCGCCCGCTTG 891

RESULT 2
BB643723 691 bp mRNA linear EST 26-OCT-2001
BB643723 RIKEN full-length enriched, adult male corpora
quadrigenina Mus musculus cDNA clone B230105C15 5', mRNA sequence.
BB643723
BB643723.1 GI:16478422
Mus musculus (house mouse)
Mus musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 691)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinegawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source
Location/Qualifiers
1..691
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B230105C15"
/sex="male"
/tissue_type="corpora quadrigenina"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male corpora
quadrigenina"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTITTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATAATTAATCCCGCCCGCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

ORIGIN
Query Match 73.3%; Score 22; DB 2; Length 691;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCTTGCCCGCCCGCTTGATCC 30
|||||
Db 552 CTGGATCTTGCCCGCCCGCTTGGAATCC 581

RESULT 3
CK840615/c 601 bp mRNA linear EST 05-MAR-2004
LOCUS
DEFINITION
UI-R-AFL-aar-c-02-0-UI-s10 UI-R-AFL Rattus norvegicus cDNA clone
UI-R-AFL-aar-c-02-0-UI 3', mRNA sequence.
ACCESSION
CK840615
VERSION
CK840615.1 GI:45193600
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 601)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel.: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/rat.html>
 The following repetitive elements were found in this cDNA
 sequence: 390-426, xMST-INTERNAL#LTR/MaLR (matched complement)
 Seq primer: ML3 FORWARD
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..601
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AF1-aar-c-02-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-AF1"
 /note="Vector: pTT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-AF1
 library is a normalized library constructed from 15 dpc
 rat atriioventricular (AV) canal. The tag is a string of 5
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6:
 791-806, 1996. Tissue provided by Jim Lin, Department of
 Biology, University of Iowa.
 TAG_TISSUE=AV canal at 15 dpc
 TAG_LIB=UI-R-AF1
 TAG_SEQ=GNAGG"

ORIGIN

Query Match 72.0%; Score 21.6; DB 7; Length 601;
 Best Local Similarity 85.7%; Pred. No. 4.4e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGGATCTCTGCGCCGCCCTTGAATTC 28

Db 509 CTGGATCTCTGCGCCGCCCTTGAATTC 482

RESULT 4

CE112892/c

LOCUS tigr-gss-dog-17000324887745 Dog Library Canis familiaris genomic,
 DEFINITION genomic survey sequence.

ACCESSION CE112892.1 GI:35179777

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 695)
 Kirkness, B.F., Baina, V., Halpern, A.L., Levy, S., Remington, K.,
 Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
 Venter, J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

MEDLINE 22875432

PUBMED 14512627

COMMENT

The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirknes@tigr.org

Class: shotgun.

FEATURES

source

Location/Qualifiers
 1..695
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"

ORIGIN

Query Match 72.0%; Score 21.6; DB 9; Length 695;
 Best Local Similarity 85.7%; Pred. No. 4.4e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GGATCCTTGCCCGCCCTTGAATTC 30

Db 427 GGATCCTTGCCCGCCCTTGAATTC 400

RESULT 5

BQ953457/c

LOCUS BQ953457 1164 bp mRNA linear EST 21-AUG-2002
 DEFINITION AGENCOURT 8818345 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6390607
 5', mRNA sequence.

ACCESSION BQ953457

VERSION BQ953457.1 GI:22368935

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1164)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13877 Row: i column: 08

High quality sequence start: 136

High quality sequence stop: 432.

FEATURES

source

1..1164

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:6390607"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_129"

/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;

Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.

Primer: Oligo dT. Average insert size 2.2 kb. Constructed

by ResGen, Invitrogen Corp. Note: this is a NIH_MGC

Library."

ORIGIN

Query Match 72.0%; Score 21.6; DB 5; Length 1164;
 Best Local Similarity 85.7%; Pred. No. 4.6e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GGATCCTTGCCCGCCCTTGAATTC 30

Db 28 GGATCCTTGCCCGCCCTTGAATTC 1

RESULT 6

BB593516/c

LOCUS BB593516 207 bp mRNA linear EST 30-NOV-2000

DEFINITION BB593516 RIKEN full-length enriched, 4 days neonate male adipose

Mus musculus cDNA clone B430101A03 5', mRNA sequence.

ACCESSION BB593516

VERSION BB593516.1 GI:11490118

KEYWORDS SOURCE ORGANISM	EST. Mus musculus (house mouse)		Best Local Similarity 88.5%; Pred.-No. 5.7e+02; Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 207)		3 GGATCCTTGGCCCGCCCTTGAATTC 28 193 GGATCCTTGGCCCGCCCTTCACTTC 168	
REFERENCE AUTHORS	Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodoyama, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toya, T., Watahiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M., and Hayashizaki, Y. RIKEN Mouse ESTs (Aizawa, K. et al. 2000)		BB753706 507 bp mRNA linear EST 17-OCT-2001 BB753706 RIKEN full-length enriched, melanocyte Mus musculus CDNA Clone G270034M03 3', mRNA sequence. BB753706 BB753706 EST. Mus musculus (house mouse)	
TITLE JOURNAL COMMENT	Unpublished (2000) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Kiteunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 507) Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001) Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.	
FEATURES source	Location/Qualifiers 1..507 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="G270034M03" /cell_type="melanocyte" /clone_lib="RIKEN full-length enriched, melanocyte"		FEATURES source 1..507 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="G270034M03" /cell_type="melanocyte" /clone_lib="RIKEN full-length enriched, melanocyte"	
ORIGIN	Query Match 70.7%; Score 21.2; DB 2; Length 207;		Query Match 70.0%; Score 21; DB 2; Length 507;	

Best Local Similarity 82.8%; Pred. No. 7.4e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TGGATCCTTCCCGCCCGCCCTTGAATTCCTC 30
DB 201 TGGTTCCTTCCCGCCCGCCCTTGAATTCAC 229

RESULT 8
BY412408
LOCUS BY412408 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus
DEFINITION cDNA clone I730090F04 3', mRNA sequence.
ACCESSION BY412408
VERSION BY412408.1 GI:26641994
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 540)
Okazaki, I., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustigich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, A., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sueniro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source 1..540
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="DBA/2"
/db_xref="taxon:10090"
/clone="I730090F04"
/cell_lines="CRL-1722 L5178Y-R"
/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

ORIGIN
Query Match 70.0%; Score 21; DB 5; Length 540;
Best Local Similarity 82.8%; Pred. No. 7.5e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 TGGATCCTTCCCGCCCGCCCTTGAATTCCTC 30
DB 244 TGGTTCCTTCCCGCCCGCCCTTGAATTCCTC 272

RESULT 9
CK773708/c
CK773708
LOCUS 962382 MARC 2BOV Bos taurus cDNA 3', mRNA sequence.
DEFINITION CK773708
ACCESSION CK773708.1 GI:42727851
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 757)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,
Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F.,
Quackenbush, J. and Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
Trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: 76 row: 1 column: 17
Seq primer: GTAATACGACCTCACTATAGG.

FEATURES
source 1..757
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"

TITLE
JOURNAL MEDLINE
PUBMED
COMMENT

```

/clone lib="MARC 2BOV"
/notes="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN
Query Match          70.0%; Score 21; DB 7; Length 757;
Best Local Similarity 82.8%; Pred. No. 7.7e+02;
Matches 24; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGATCTTGGCCCGCCCTTGAATCC 29
    |||:|||||:|||||:|||||:|||||
Db 636 CTGAAGCCTTGTCGCCCGCCCTTGTATGCC 608

RESULT 10
AL567601/c
LOCUS
DEFINITION
AL567601 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CSODF030YH23 3-PRIME, mRNA sequence.
ACCESSION
AL567601
VERSION
AL567601.3 GI:46234082
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 1139)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
On Feb 16, 2001 this sequence version replaced gi:31290473.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 2863.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CSODF030CD12NP1&c=2863.f.

FEATURES
Location/Qualifiers
source
1..1139
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF030YH23"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match          70.0%; Score 21; DB 1; Length 1139;
Best Local Similarity 72.0%; Pred. No. 8.1e+02;
Matches 18; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 6 TCCTTGGCCCGCCCTTGAATCC 30
    |||:|||||:|||||:|||||:|||||
Db 1042 TCCTKSCCCCGCCCTTGMAGTSCC 1018

RESULT 11
CD497293
LOCUS
DEFINITION
CDA27-A09 y1d-s SHGC-CDA Gasterosteus aculeatus cDNA clone
CDA27-A09 3', mRNA sequence.
CD497293
EST..
CD497293.1 GI:31424324
Gasterosteus aculeatus (three spined stickleback)
Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthopterygia; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1158)
Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 27
High quality sequence start: 43
High quality sequence stop: 561.
Location/Qualifiers
source
1..1158
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA27-A09"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"

ORIGIN
Query Match          69.3%; Score 20.8; DB 6; Length 1158;
Best Local Similarity 91.7%; Pred. No. 9.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGATCTTGCCCGCCCGCCCTTGAAT 26
    |||:|||||:|||||:|||||:|||||
Db 1134 GGATCTTGCCCGCCCGCCCTTTAAT 1157

RESULT 12
CN704387/c
LOCUS
DEFINITION
CN704387 E0487A12-5 NTA Mouse E11.5 whole embryo cDNA library (long) Mus
musculus cDNA clone NIA:E0487A12 IMAGE:30876875 5', mRNA sequence.
ACCESSION
CN704387
VERSION
CN704387.1 GI:47473136
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 565)
Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y.,
VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Baesey,U.C.,
Wang,Y., Carter,M.G., Hamatani,T., Aiba,K., Akutsu,H., Sharova,L.,
Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S.,
Nagaraja,R., Boheler,K.R., Taub,D., Hodes,R.J., Longo,D.L.,
Schlessinger,D., Keller,J., Klotz,E., Kelseo,G., Umezawa,A.,
Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A.,
D'Urso,M., Kelseo,J., Hide,W. and Ko,M.S.
Transcriptome analysis of mouse stem cells and early embryos
PLOS Biol. 1 (3), 410-419 (2003)
Contact: dawood.b.dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: E0487 row: A column: 12
Seq primer: M13 Reverse
High quality sequence stop: 565
POLYA=No.

TITLE
JOURNAL
COMMENT

FEATURES
source
1..565
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:E0487A12-5"
/db_xref="taxon:10090"
/clone="NIA:E0487A12 IMAGE:30876875"
/tissue_type="whole embryo including extraembryonic
tissues at 11.5-days postcoitum"
/dev_stage="E11.5"
/lab_host="DH10B"
/clone_lib="NIA Mouse E11.5 whole embryo cDNA library
(long)"
/notes="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 3 embryos at 11.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
Oligo(dT) primer [Invitrogen:
5'-pGACTAGTCTAGATCGGCGCGCCCTTTT-3'] from
2ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 3.3Kb. The library was
constructed by Yulan Piao."

ORIGIN
Query Match 68.7%; Score 20.6; DB 7; Length 565;
Best Local Similarity 85.2%; Pred. No. 1.1e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GATCCTTGGCCCGCCCTTGAATCC 30
|||||
Db 507 GATCTTTGCCCGCTCGTGAATCC 481
|||||

RESULT 13
CN698326/c
LOCUS
DEFINITION E0406C01-5 NIA Mouse E11.5 whole embryo cDNA library (Long) Mus

```

```

musculus cDNA clone NIA:E0406C01 IMAGE:30869112 5', mRNA sequence.
CN698326
CN698326.1 GI:47467075
EST.
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 654)
Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y.,
VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Baesey,U.C.,
Wang,Y., Carter,M.G., Hamatani,T., Aiba,K., Akutsu,H., Sharova,L.,
Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S.,
Nagaraja,R., Boheler,K.R., Taub,D., Hodes,R.J., Longo,D.L.,
Schlessinger,D., Keller,J., Klotz,E., Kelseo,G., Umezawa,A.,
Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A.,
D'Urso,M., Kelseo,J., Hide,W. and Ko,M.S.
Transcriptome analysis of mouse stem cells and early embryos
PLOS Biol. 1 (3), 410-419 (2003)
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: E0406 row: C column: 01
Seq primer: M13 Reverse
High quality sequence stop: 654
POLYA=No.

FEATURES
source
1..654
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:E0406C01-5"
/db_xref="taxon:10090"
/clone="NIA:E0406C01 IMAGE:30869112"
/tissue_type="whole embryo including extraembryonic
tissues at 11.5-days postcoitum"
/dev_stage="E11.5"
/lab_host="DH10B"
/clone_lib="NIA Mouse E11.5 whole embryo cDNA library
(long)"
/notes="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 3 embryos at 11.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
Oligo(dT) primer [Invitrogen:
5'-pGACTAGTCTAGATCGGCGCGCCCTTTT-3'] from
2ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 3.3Kb. The library was
constructed by Yulan Piao."

ORIGIN
Query Match 68.7%; Score 20.6; DB 7; Length 654;
Best Local Similarity 85.2%; Pred. No. 1.1e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GATCCTTGGCCCGCCCTTGAATCC 30
|||||

```

Db 507 GATCTTGGCCCGCTGGTGAATTC 481

RESULT 14
AK039759/c
LOCUS
DEFINITION
AK039759 3697 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male spinal cord cDNA, RIKEN full-length
enriched library, clone:A330104K03 product:Ellis van Creveld gene
homolog (human), full insert sequence.

ACCESSION
AK039759
VERSION
AK039759.1 GI:26087386
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913
PUBMED
11076861

REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
AUTHORS

REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
AUTHORS

REFERENCE
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, H., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp)

COMMENT
URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>

FEATURES
Location/Qualifiers
1..3697
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:A330104K03"
/db_xref="taxon:10090"
/clone="A330104K03"
/sex="male"
/tissue_type="spinal cord"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..3697
/note="Ellis van Creveld gene homolog (human)
(MGDI|MG1:1890596, GB|NM_021292, evidence: BLASTN, 99%,
match=2739)"

ORIGIN
Query Match 68.7%; Score 20.6; DB 3; Length 3697;
Best Local Similarity 85.2%; Pred. No. 1.3e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
4 GATCTTGGCCCGCTGGTGAATTC 30
|||||
1397 GATCTTGGCCCGCTGGTGAATTC 1371

RESULT 15
AA763854
LOCUS
DEFINITION
IMAGE:1225201 5', mRNA sequence.
ACCESSION
AA763854
VERSION
AA763854.1 GI:2811376
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 378)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:650793
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 360.
Location/Qualifiers
1..378
/organism="Mus musculus"
/mol_type="mRNA"


```

/clone_lib="normal cattle brain"
/notes="Organ: Brain; Vector: pSport1; Site 1: NotI;
Site 2: SalI; Random primed and directionally cloned in
pSport1 vector using NotI
(5'-pGACTAGTCTTAGATCGGCGGCC (T)15-3' and SalI 5'-
TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"

ORIGIN

Query Match      68.0%; Score 20.4; DB 7; Length 594;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGATCCTTGCCCGCCCGCCCTTGAATCC 30
    ||||| ||||| ||||| ||||| |||||
Db 286 CTGACCCCTTGCCCGCCCGCCCTTGAGACCC 315

RESULT 22
CE337240      691 bp DNA linear GSS 26-SEP-2003
LOCUS      tigr-gss-dog-17000334049010 Dog Library Canis familiaris genomic,
DEFINITION      genomic survey sequence.
ACCESSION      CE337240
VERSION      CE337240.1 GI:36158806
KEYWORDS      GSS.
SOURCE      Canis familiaris (dog)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE      1 (bases 1 to 691)
AUTHORS      Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K.,
      Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
      Venter,J.C.
TITLE      The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED      14512627
COMMENT      Contact: Kirkness EF
      The Institute for Genomic Research
      Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
      Rockville, MD 20850, USA
      Tel: 301-838-0200
      Fax: 301-838-0208
      Email: ekirknes@tigr.org
      Class: shotgun.
FEATURES
      source
      Location/Qualifiers
          1..691
              /organism="Canis familiaris"
              /mol_type="genomic DNA"
              /strain="Standard Poodle"
              /db_xref="taxon:9615"
              /clone_lib="Dog Library"
              /note="Site 1: BstXI; Libraries were prepared from
              peripheral blood"

ORIGIN

Query Match      68.0%; Score 20.4; DB 9; Length 691;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGATCCTTGCCCGCCCGCCCTTGAATCC 30
    ||||| ||||| ||||| ||||| |||||
Db 160 CTTATCCTTTCCTGCTGCTTGAATCC 189

RESULT 23
B1555868      724 bp mRNA linear EST 05-SEP-2001
LOCUS      603238173F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5291035 5',
DEFINITION      mRNA sequence.
ACCESSION      B1555868
VERSION      B1555868.1 GI:15443182
KEYWORDS      EST.

```

```

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 724)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1736 row: a column: 20
High quality sequence stop: 716.
FEATURES
      source
      Location/Qualifiers
          1..724
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="129, C57BL/6J, FVB/N"
              /db_xref="taxon:10090"
              /clone="IMAGE:5291035"
              /tissue type="tumor, gross tissue"
              /dev stage="10 months"
              /lab_host="DH108"
              /clone_lib="NCI_CGAP Mam3"
              /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
              Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
              Library constructed by Life Technologies. Investigators
              providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
              Reference for transgenic model: Xu et al., Nature Genetics
              22, 37-43 (1999)."
ORIGIN

Query Match      68.0%; Score 20.4; DB 4; Length 724;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCGCCCTTGAATCC 30
    ||||| ||||| ||||| ||||| |||||
Db 138 CTGGACCCGCGCCCTCCCGGTAATCC 167

RESULT 24
BC028653      725 bp mRNA linear HTC 20-SEP-2002
LOCUS      Mus musculus, Similar to LOC169382, clone IMAGE:1225201, mRNA.
DEFINITION
ACCESSION      BC028653
VERSION      BC028653.1 GI:20306565
KEYWORDS      HTC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 725)
Strausberg,R.
Direct Submission
Submitted (23-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
CDNA Library Preparation: Soares Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org

```

contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 66 Row: d Column: 8
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..725
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1225201"
/tissue_type="Thymus gland, mouse"
/clone_lib="Soares_thymus_2NbMT"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac"

ORIGIN

Query Match 68.0%; Score 20.4; DB 3; Length 725;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
|||||
DB 271 CTGGAGCCCTGGCCCTCTCTTGAATTCCC 300

RESULT 25

AG441314
LOCUS AG441314 733 bp DNA linear GSS 03-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-322K24.T7, genomic survey sequence.

ACCESSION AG441314

VERSION AG441314.1

KEYWORDS GI:48084377

SOURCE GSS.

ORGANISM Mus musculus molossinus

Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

BAC end Sequences of Library MSMg01

Unpublished

2 (bases 1 to 733)

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.go.jp/;

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Tsukuba Institute, Bio Resource Center,

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : T7

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1..733

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

FEATURES

Location/Qualifiers
1..733
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"

ORIGIN

Query Match 68.0%; Score 20.4; DB 9; Length 733;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
|||||
DB 631 CTGGATCCTTGGCCCGCCCTTGGCCATCCC 660

RESULT 26

AG413802
LOCUS AG413802 747 bp DNA linear GSS 03-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-275E24.TJ, genomic survey sequence.

ACCESSION AG413802

VERSION AG413802.1

KEYWORDS GI:48056590

SOURCE GSS.

ORGANISM Mus musculus molossinus

Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

BAC end Sequences of Library MSMg01

Unpublished

2 (bases 1 to 747)

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.go.jp/;

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Tsukuba Institute, Bio Resource Center,

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1..747

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSMg01-275E24.TJ"

/sex="male"

/tissue_type="mixture of kidney and spleen"

/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 68.0%; Score 20.4; DB 9; Length 747;
Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTT 22
|||||
DB 686 CTGGATCTGGCCCGCCCTT 707

```

RESULT 27
CR201952      753 bp      DNA      linear      GSS 06-JUL-2004
LOCUS
DEFINITION   Chromosome engineering clone MHPN334b07, genomic survey sequence.
ACCESSION   CR201952
VERSION      GSS; genome survey sequence; MICR.
KEYWORDS     Mus musculus (house mouse)
SOURCE
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 753)
AUTHORS      Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
              Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
              Rogers,J. and Bradley,A.
TITLE        Direct Submission
JOURNAL      Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. http://www.sanger.ac.uk/MICR
FEATURES     Location/Qualifiers
              source
                1..753
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /clone="MHPN334b07"
                /clone_lib="MHPN"
ORIGIN
Query Match      68.0%; Score 20.4; DB 9; Length 753;
Best Local Similarity 80.0%; Pred. No. 1.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 213 CTGGATCCTTCCCTCCCGTGGCATCCC 242

RESULT 28
CR124383/c   833 bp      DNA      linear      GSS 06-JUL-2004
LOCUS
DEFINITION   Forward strand read from insert in 3'HPRT insertion targeting and
              chromosome engineering clone MHP407a20, genomic survey sequence.
ACCESSION   CR124383
VERSION      GSS; genome survey sequence; MICR.
KEYWORDS     Mus musculus (house mouse)
SOURCE
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 833)
AUTHORS      Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
              Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
              Rogers,J. and Bradley,A.
TITLE        Direct Submission
JOURNAL      Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. http://www.sanger.ac.uk/MICR
FEATURES     Location/Qualifiers
              source
                1..833
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /clone="MHP407a20"
                /clone_lib="MHPN"
ORIGIN
Query Match      68.0%; Score 20.4; DB 9; Length 833;
Best Local Similarity 80.0%; Pred. No. 1.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 78 CTGCTTCCTTGCACACTACCCCTTGAATCAC 49

RESULT 29
AG082767/c   1039 bp     DNA      linear      GSS 03-NOV-2001
LOCUS
DEFINITION   Pan troglodytes DNA, clone: PTB-079M04.F, genomic survey sequence.
ACCESSION   AG082767
VERSION      AG082767.1 GI:16634569
KEYWORDS     Pan troglodytes (chimpanzee)
SOURCE
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE    1
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
              Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1039)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
              Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:chimpes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
              Clones are derived from the chimpanzee BAC library PTB This BAC end
              was generated during the R&D process and may have higher chance of
              clone tracking errors.
COMMENT
PRIMERS
Sequencing: -21M13
LIBRARY
Vector          : pKS145
R.Site 1        : SacI
R.Site 2        : SacI.
FEATURES     Location/Qualifiers
              source
                1..1039
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                /clone="PTB-079M04.F"
                /sex="male"
                /cell_type="lymphoblast"
                /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      68.0%; Score 20.4; DB 9; Length 1039;
Best Local Similarity 80.0%; Pred. No. 1.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 455 CTGATTCCTTCTCCGCCCATTTGACTCCC 426

RESULT 30
AI646086/c   340 bp     mRNA      linear      EST 29-APR-1999
LOCUS
DEFINITION   IMAGE:1247666 3', mRNA sequence.
ACCESSION   AI646086
VERSION      AI646086.1 GI:4724561
KEYWORDS     EST.
SOURCE
ORGANISM     Mus musculus (house mouse)
              Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 340)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
              National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
              Contact: Robert Strausberg, Ph.D.
              
```

Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:661354
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end.

FEATURES
 source
 1..340
 /location=Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:1247666"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH108"
 /clone_lib="Soares mammary_gland_NMLMG"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaído."

ORIGIN

Query Match 66.7%; Score 20; DB 1; Length 340;
 Best Local Similarity 82.1%; Pred. No. 1.8e+03;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GGATCCTTCCCGCCCGCCCTTGAATCCC 30
 |||||
 Db 340 GGTTCCTCCCTGCGCCCTTGAATCCAC 313

RESULT 31

W61990
 LOCUS
 DEFINITION
 m83b02.r1 Soares mouse embryo NDM13.5 14.5 Mus musculus cDNA
 clone IMAGE:374955 5', mRNA sequence.

W61990
 VERSION
 KEYWORDS
 SOURCE
 Mus musculus (house mouse)
 ORGANISM

REFERENCE
 AUTHORS
 1 (bases 1 to 373)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

TITLE
 JOURNAL
 COMMENT
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:236387
 Seq primer: mob.REGA+ET
 High quality sequence stop: 352.
 Location/Qualifiers

FEATURES

source
 1..373
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

/clone="IMAGE:374955"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH108"
 /clone_lib="Soares mouse embryo NDM13.5 14.5"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGTGGAGCGCGCCCGGAATTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M. Fatima Bonaído."

ORIGIN

Query Match 66.7%; Score 20; DB 7; Length 373;
 Best Local Similarity 82.1%; Pred. No. 1.8e+03;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTCCCGCCCGCCCTTGAATTC 28
 |||||
 Db 48 CTGGATCCTTCCCGCCCAACCCCTGAGTTC 75

RESULT 32

W75418
 LOCUS
 DEFINITION
 me50b02.r1 Soares mouse embryo NDM13.5 14.5 Mus musculus cDNA
 clone IMAGE:390891 5', mRNA sequence.

W75418
 VERSION
 KEYWORDS
 SOURCE
 Mus musculus (house mouse)
 ORGANISM

REFERENCE
 AUTHORS
 1 (bases 1 to 417)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

TITLE
 JOURNAL
 COMMENT
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:242723
 Seq primer: mob.REGA+ET
 High quality sequence stop: 352.
 Location/Qualifiers

FEATURES

source
 1..417
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:390891"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH108"
 /clone_lib="Soares mouse embryo NDM13.5 14.5"

```
FEATURES
source
Location/Qualifiers
1..511
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A930102E24"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-l
```

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory, RIKEN, Division of Experimental Animal Research in RIKEN contributed to prepare mouse tissues. 1st strand cDNA was prepared with a primer [5', GAGAGAGAGAGAGCTCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATCTCGATTATTAATTAATTCCTCCCCCCCC-3']. cDNA was cleaved with *XhoI* and *BamHI*. Vector: a modified pBluescript KS(+) after bulk excision from LFLC 1. Retina RNA was provided by Stefano Gustinich Department of Neurobiology, Harvard Medical School, 2 Longwood Ave., Boston, MA02115, USA, whose assistance gratefully acknowledge."

Prepared by: Seung-Hoon Lee, Department of Cell Biology, Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was prepared with a primer [5', GAGAGAGAAAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was prepared by using trahalose thermo-activated reverse

to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGGAGAGATCTCGATTAAATTAATCCCCCCCCC-3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(-) after bulk excision from Lambda ϕ 11. Retina RNA was provided by Stefano Gustinich, FLC I.-Retina Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge.

Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

ORIGIN					
Query Match 66.7%; Score 20; DB 2; Length 511;					
Best Local Similarity 82.1%; Pred. No. 1.9e+03;					
Matches	23;	Conservative	0;	Mismatches	5; Indels 0; Gaps 0;
QY	1	CTGGATCCTTGGCCCCGCCCTTGAAATTC	28		
Dd	19	CTGGATCCTTGGCCCCAACCCTGAGTTC	46		
RESULT 34					
BG067160/c					
LOCUS					
BG067160					
H3051C07-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone					
DEFINITION					
H3051C07 3'. mRNA sequence.					
EST 17-DEC-2003					

EST.	07.100.2	GI:40013002
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 517)	
AUTHORS	Tanaka T.S., Jaradat S.A., Lim M.K., Kargul G.J., Wang X., Grabovac M.J., Pantano S., Sano Y., Piao Y., Nagaraja R., Doi H., Wood W.H. III, Becker K.G. and Ko M.S.H.	
TITLE	Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)	
MEDLINE	20381348	
PUBLISHED	10922068	
COMMENT	On Jan 26, 2001 this sequence version replaced gi:12549729. Other ESTs: H3051C07-5 Contact: George J. Kargul Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@lgsum.grc.nia.nih.gov This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.	

Plate: H3051 row: C column: 07
 Seq primer: -21M13 Forward
 High quality sequence stop: 517
 POLYA=Yes.

FEATURES

source
 1. 517
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="niaEST:H3051C07-3"
 /db_xref="taxon:10090"
 /clone="H3051C07"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN

Query Match 66.7%; Score 20; DB 4; Length 517;
 Best Local Similarity 82.1%; Pred. No. 1.9e+03;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCGCCCTTGAATTC 28
 |||||
 Db 490 CTGGATCCTTGCCTCCCAACCTGATTC 463

RESULT 35

CA884170 520 bp mRNA linear EST 20-DEC-2002
 B0108D11-3N NIA Mouse Neural Stem Cell (Differentiated) cDNA
 Library (Long) Mus musculus cDNA clone NIA:B0108D11 IMAGE:30094606
 3', mRNA sequence.

ACCESSION CA884170
 VERSION CA884170
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 520)
 AUTHORS Piao, Y., Dudekula, D.B., Qian, Y., Martin, P.R., Aiba, K., Vescovi, A.L. and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated) cDNA Library (Long)
 JOURNAL Unpublished (2002)
 COMMENT Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@gsun.grc.nia.nih.gov
 Plate: B0108 row: D column: 11
 Seq primer: -21M13 Forward

High quality sequence stop: 520
 POLYA=Yes.

FEATURES

source
 1. 520
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CD1"
 /db_xref="niaEST:B0108D11-3N"
 /db_xref="taxon:10090"
 /clone="NIA:B0108D11 IMAGE:30094606"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
 cDNA Library (Long)"
 /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
 NotI; Mouse cDNA project by the Laboratory of Genetics,
 National Institute on Aging (NIA), Intramural Research
 Program, NIH (http://gsun.grc.nia.nih.gov/cDNA). This is
 a long-transcript enriched cDNA library (Ref. Genome Res.
 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
 obtained from Dr. Angelo L. Vescovi (Institute for Stem
 Cell Research, Italy). Double-stranded cDNAs were
 synthesized with an Oligo(dT) primer [Invitrogen:
 5'-PGACTAGTCTAGATCGAGCGCCCTTTT-3'] from
 2.0 Microgram of total RNA, treated with T4 DNA
 polymerase, and purified by ethanol-precipitation. The
 cDNAs were ligated to Lone-linker LL-Sal4, purified by
 phenol/chloroform, and separated from free linkers by
 Centricon 100. Then, the cDNAs were amplified by
 long-range high fidelity PCR using Ex Taq polymerase
 (Takara) with a primer Sal4-S. The products were purified
 by phenol/chloroform and Centricon 100. The cDNAs were
 digested with SalI and NotI enzymes and cloned into
 SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
 coli host was transformed with the ligation mixture by the
 standard chemical method. The average insert size is about
 3.2 kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 66.7%; Score 20; DB 6; Length 520;
 Best Local Similarity 82.1%; Pred. No. 1.9e+03;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCGCCCTTGAATTC 28
 |||||
 Db 498 CTGGATCCTTGCCTCCCAACCTGATTC 471

RESULT 36

B0523276 528 bp mRNA linear EST 10-JUN-2002
 LOCUS NISC_n121c09.x1 NICHD_XGC_Emb7 Xenopus tropicalis cDNA clone
 DEFINITION IMAGE:5379448 3', mRNA sequence.

ACCESSION B0523276
 VERSION B0523276
 KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 528)
 AUTHORS NIH-XGC http://image.llnl.gov/image/html/xenopuslib.info.shtml.
 TITLE National Institute of Child Health and Human Development, National
 Cancer Institute, Xenopus Gene Collection
 JOURNAL Unpublished (2002)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 cDNA Library Preparation:
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CGAP clone distribution information can be


```

RESULT 39
A2744176/c
LOCUS
DEFINITION
  RPCI-24-146F1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-146F1,
  genomic survey sequence.
ACCESSION
  A2744176
VERSION
  A2744176.1 GI:12524768
KEYWORDS
  GSS.
SOURCE
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 661)
  Tseng, G., Geer, K., Krol, M., Shvartsbeyn, A., Akinret, B., Levins, M.,
  Russell, D., de Jong, P. and Fraser, C.M.
  Mouse BAC End Sequences from Library RPCI-24
  Unpublished (1999)
  Other GSSs: RPCI-24-146F1.TV
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-24. For BAC
  library availability, please contact Pieter de Jong
  (pdejong@mail.cho.org). Clones may be purchased from BACPAC
  Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
  plate: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
  Plate: 146 row: F column: 1
  Seq primer: SP6
  Class: BAC ends.
  Location/Qualifiers
    1..661
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPCI-24-146F1"
      /sex="Male"
      /cell_type="Spleen/Brain"
      /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
      RPCI-24 Mouse BAC library produced by Pieter de Jong. The
      library was cloned in the pTARBAC1 cloning vector at the
      BamHI sites using MboI partially digested male C57BL/6J
      DNA."
  ORIGIN
    Query Match 66.7%; Score 20; DB 8; Length 661;
    Best Local Similarity 82.1%; Pred. No. 1.9e+03;
    Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GGATCCTTGGCCCGCCCTTGAATCCC 30
Db 652 GGATCCTTGGCCCTTGGCCCTTGAATCCC 625

RESULT 40
BY757487
LOCUS
DEFINITION
  BY757487 RIKEN full-length enriched, 15 days pregnant adult female
  placenta Mus musculus cDNA clone I530017K15 3', mRNA sequence.
ACCESSION
  BY757487
VERSION
  BY757487.1 GI:27190993
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 674)

```

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Fraser, K.S., Gaasterland, T.,
 Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shmida, M.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

MEDLINE

22354683

PUBMED

12466851

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
 Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
 Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
 Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
 Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
 Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
 Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES

Location/Qualifiers
 1..674
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I530017K15"

```

/sex="female"
/tissue_type="placenta"
/dev_stage="15 days pregnant adult"
/clone_lib="RIKEN full-length enriched, 15 days pregnant
adult female placenta"

ORIGIN
Query Match 66.7%; Score 20; DB 6; Length 674;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
Db 185 CTGGATCCTTGGCCCAACCTGAGTTC 212

RESULT 41
CR085351
LOCUS CR085351 680 bp DNA linear GSS 05-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHP262e24, genomic survey sequence.
ACCESSION CR085351
VERSION CR085351.1 GI:49818942
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 680)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submision
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES
source
1..680
/mol_type="Mus musculus"
/db_xref="genomic DNA"
/clone_lib="MHP262e24"
/clone_lib="MHP2"

ORIGIN
Query Match 66.7%; Score 20; DB 9; Length 680;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
Db 2 CCGGATCCTTGGCTCACCCTTTAATTC 29

RESULT 42
AW187871
LOCUS AW187871 684 bp mRNA linear EST 30-NOV-1999
DEFINITION BNLGH13430 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
to cobalamin-independent methionine synthase, mRNA sequence.
ACCESSION AW187871
VERSION AW187871.1 GI:6462307
KEYWORDS EST.
SOURCE Gossypium hirsutum (upland cotton)
ORGANISM Gossypium hirsutum

REFERENCE
1 (bases 1 to 684)
AUTHORS Blewitt,M., Matz,E.C. and Burr,B.
TITLE ESTs from developing cotton fiber (1999b)
JOURNAL Unpublished (1999)
COMMENT Contact: Ben Burr
Biology Department
Brookhaven National Laboratory

/sex="female"
/tissue_type="placenta"
/dev_stage="15 days pregnant adult"
/clone_lib="RIKEN full-length enriched, 15 days pregnant
adult female placenta"

ORIGIN
Query Match 66.7%; Score 20; DB 6; Length 674;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
Db 185 CTGGATCCTTGGCCCAACCTGAGTTC 212

RESULT 41
CR085351
LOCUS CR085351 680 bp DNA linear GSS 05-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHP262e24, genomic survey sequence.
ACCESSION CR085351
VERSION CR085351.1 GI:49818942
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 680)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submision
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES
source
1..680
/mol_type="Mus musculus"
/db_xref="genomic DNA"
/clone_lib="MHP262e24"
/clone_lib="MHP2"

ORIGIN
Query Match 66.7%; Score 20; DB 9; Length 680;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
Db 2 CCGGATCCTTGGCTCACCCTTTAATTC 29

RESULT 42
AW187871
LOCUS AW187871 684 bp mRNA linear EST 30-NOV-1999
DEFINITION BNLGH13430 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
to cobalamin-independent methionine synthase, mRNA sequence.
ACCESSION AW187871
VERSION AW187871.1 GI:6462307
KEYWORDS EST.
SOURCE Gossypium hirsutum (upland cotton)
ORGANISM Gossypium hirsutum

REFERENCE
1 (bases 1 to 684)
AUTHORS Blewitt,M., Matz,E.C. and Burr,B.
TITLE ESTs from developing cotton fiber (1999b)
JOURNAL Unpublished (1999)
COMMENT Contact: Ben Burr
Biology Department
Brookhaven National Laboratory

Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@nrlx1.bnl.gov
Seq primer: T3 Primer.
Location/Qualifiers
1..684
/organism="Gossypium hirsutum"
/mol_type="mRNA"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/clone_lib="Six-day Cotton fiber"
/note="Vector: pBluescript II KS+"

FEATURES
source
1..684
Location/Qualifiers
1..684
/mol_type="Mus musculus"
/db_xref="genomic DNA"
/clone_lib="MHP262e24"
/clone_lib="MHP2"

ORIGIN
Query Match 66.7%; Score 20; DB 2; Length 684;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 29
    |||||
Db 536 TGGATCCTTGGCCCGCCCTTGAATTC 563

RESULT 43
CR047114
LOCUS CR047114 688 bp DNA linear GSS 05-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHP88m16, genomic survey sequence.
ACCESSION CR047114
VERSION CR047114.1 GI:49780172
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 688)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submision
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES
source
1..688
/mol_type="Mus musculus"
/db_xref="genomic DNA"
/clone_lib="MHP88m16"
/clone_lib="MHP2"

ORIGIN
Query Match 66.7%; Score 20; DB 9; Length 688;
Best Local Similarity 82.1%; Pred. No. 1.9e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
Db 578 CTGGATCCTTGGCCCAACCTGAGTTC 605

RESULT 44
CG436661
LOCUS CG436661 889 bp DNA linear GSS 17-SEP-2003
DEFINITION OGTCG93TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA0861P17,
genomic survey sequence.
ACCESSION CG436661
VERSION CG436661.1 GI:34813200
KEYWORDS GSS.
SOURCE Zea mays

```

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 889)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Contact: Cathy Whitelaw
TIGR

REFERENCE
AUTHORSTITLE
JOURNAL
COMMENT

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..889
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMWBMA0861P17"
/clone_lib="ZM 0.7 1.5 kb"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 889;
Best Local Similarity 82.1%; Pred. No. 2e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY

3 GGATCCTTGCCTGCGCCCTTGATTC 30

DB

279 GGATTCCTGCGCCCGCCATGATCCTC 252

RESULT 45

B1525356

LOCUS

602924465F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5056793 5',
mRNA sequence.

ACCESSION

B1525356

VERSION

B1525356.1 GI:15350148

KEYWORDS

EST.

SOURCE

Mus musculus
(house mouse)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1154 row: a column: 18
High quality sequence start: 24
High quality sequence stop: 745.

FEATURES

source

Location/Qualifiers

1..890
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C25CH II"

/db_xref="taxon:10090"

/clone="IMAGE:5056793"

/tissue type="pooled lung tumors"

/lab host="DH10B (phage-resistant)"

/clone_lib="NCI_CGAP_Lu33"

/note="Organ: lung; Vector: pTT3D-Pac (Pharmacia) with a
modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCTCTGTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pTT3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 66.7%; Score 20; DB 4; Length 890;
Best Local Similarity 82.1%; Pred. No. 2e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY

1 CTGGATCCTTGCCTGCGCCCTTGATTC 28

DB

380 CTGGATCCTTGCCTGCGCCCAACCTGATTC 407

RESULT 46

CL126864

LOCUS

CL126864 980 bp DNA linear GSS 05-JAN-2004
ISB1-91117 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-91117,
genomic survey sequence.

ACCESSION

CL126864

VERSION

CL126864.1 GI:40620499

KEYWORDS

GSS.

SOURCE

Xenopus tropicalis (western clawed frog)

ORGANISM

REFERENCE

AUTHORS

Krenitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E., and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)

TITLE

JOURNAL

COMMENT

Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu

Insert Length: 75000 Std Error: 0.00

Seq primer: T7 TAATACGACTCACTATAGG

Class: BAC ends

High quality sequence start: 16

High quality sequence stop: 173.

FEATURES

source

Location/Qualifiers

1..980
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-91117"
/clone_lib="ISB1"
/note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 980;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 CTGGATCCTTGCCTGCGCCGCC 20

DB

78 CTGGATCCTTGCCTGCGCCGCC 97

```

RESULT 47
CF223569/c
LOCUS
DEFINITION
  CF223569 15064707 NICHG XGC Emb7 Xenopus tropicalis cDNA clone
  IMAGE:6976671 5', mRNA sequence.
ACCESSION
  CF223569
VERSION
  CF223569.1 GI:33424277
KEYWORDS
  EST.
SOURCE
  Xenopus tropicalis (western clawed frog)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
    Xenopodinae; Xenopus; Siurana.
  1 (bases 1 to 1015)
  NIH-MGC http://mgc.nhl.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Daniela S. Gerhard, Ph.D.
  Office of Cancer Genomics
  National Cancer Institute / NIH
  Bldg. 31 Rm10A07 Bethesda, MD 20892
  Email: cgabbs-r@mail.nih.gov
  Tissue Procurement: Robert M. Grainger
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM14630 Row: d Column: 14
  High quality sequence stop: 677.
  Location/Qualifiers
    1..1015
      /organism="Xenopus tropicalis"
      /mol_type="mRNA"
      /db_xref="taxon:8364"
      /clone="IMAGE:6976671"
      /tissue_type="tailbud"
      /dev_stage="embryo, stages 20-27"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NICHG XGC Emb7"
      /note="Vector: pCMV-SF076.1; Site 1: NotI; Site 2: EcoRV;
      Cloned unidirectionally. Primer: Oligo dt. Average insert
      size 2.1 kb. Constructed by invitrogen. Note: This is a
      Xenopus Gene Collection (XGC) library."
ORIGIN
  Query Match 66.7%; Score 20; DB 7; Length 1015;
  Best Local Similarity 82.1%; Pred. No. 28+03;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  QY 3 GGATCCTTGGCCCGCCCTTGAATCCC 30
  Db 918 GCATCATTCCTCCACCCCTTATATCCC 891

  RESULT 48
  CC237106/c
  LOCUS
  DEFINITION
    CC237106 1537 bp DNA linear GSS 12-MAY-2003
    CH261-191E5_Sp6.1 CH261 Gallus gallus genomic clone CH261-191E5,
    genomic survey sequence.
  ACCESSION
    CC237106
  VERSION
    CC237106.1 GI:30563769
  KEYWORDS
    GSS.
  SOURCE
    Gallus gallus (chicken)
    ORGANISM
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
      Phasianinae; Gallus.
      1 (bases 1 to 1537)
      Krenitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
      Warren,W., Graves,T., Mardis,E. and Wilson,R.
      Gallus gallus BAC End Reads
  TITLE

  JOURNAL
  COMMENT
  Unpublished (2003)
  Contact: Richard K. Wilson
  Genome Sequencing Center
  Washington University School of Medicine
  Email: submissions@wustl.edu
  Insert Length: 182000 Std Error: 0.00
  Seq primer: Sp6 ATTAGGTGACACTATAG
  Class: BAC ends
  High quality sequence start: 332
  High quality sequence stop: 1022.
  Location/Qualifiers
    1..1537
      /organism="Gallus gallus"
      /mol_type="genomic DNA"
      /strain="Red Jungle Fowl"
      /db_xref="taxon:9031"
      /clone="CH261-191E5"
      /sex="female"
      /cell_line="UCD001, inbred 256"
      /clone_lib="CH261"
      /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
      CH261 Female Chicken library - For library and clone
      ordering information: http://www.chori.org/bacpac"
ORIGIN
  Query Match 66.7%; Score 20; DB 8; Length 1537;
  Best Local Similarity 82.1%; Pred. No. 2.1e+03;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  QY 2 TCGATCCTTGGCCCGCCCTTGAATTC 29
  Db 1198 TGTGCTTCCCTCCCTCCCTTGAATTC 1171

  RESULT 49
  BX308381/c
  LOCUS
  DEFINITION
    BX308381 tcay Oncorhynchus mykiss cDNA clone tcay0020b.o.11 3prim,
    mRNA sequence.
  ACCESSION
    BX308381
  VERSION
    BX308381.2 GI:40224293
  KEYWORDS
    EST.
  SOURCE
    Oncorhynchus mykiss (rainbow trout)
    ORGANISM
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
      Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
      1 (bases 1 to 669)
      Govoroun,M., Guiguen,Y. and Le Gac,F.
      Construction and primary characterization of normalized cDNA
      libraries in rainbow trout, Oncorhynchus mykiss
      Unpublished (2003)
      On Apr 7, 2003 this sequence version replaced gi:29589026.
      Contact: Guiguen Y
      INRA - SCRIBE
      Campus de Beaulieu, RENNES cedex, 35042, France
      Tel.: 02.23.48.50.09
      Fax: 02.23.48.50.20
      Email: Yann.Guiguen@beaulieu.rennes.inra.fr
      Sequence cleaned of vector, adaptor and repetitions. Contact us
      at signeaupport@jouy.inra.fr to obtain the chromatogram of this
      sequence.
      Plate: 0020 Row: 0 Column: 11
      Seq primer: M13P.
      Location/Qualifiers
        1..669
          /organism="Oncorhynchus mykiss"
          /mol_type="mRNA"
          /db_xref="taxon:8022"
          /clone="tcay0020b.o.11"
          /tissue_type="adipose tissue, blood, brain,
          differentiating gonads, gills, interrenal, intestine,
          kidney, liver, muscle, ovary, pituitary, testis"
  TITLE
  JOURNAL
  COMMENT
  Unpublished (2003)
  Contact: Richard K. Wilson
  Genome Sequencing Center
  Washington University School of Medicine
  Email: submissions@wustl.edu
  Insert Length: 182000 Std Error: 0.00
  Seq primer: Sp6 ATTAGGTGACACTATAG
  Class: BAC ends
  High quality sequence start: 332
  High quality sequence stop: 1022.
  Location/Qualifiers
    1..1537
      /organism="Gallus gallus"
      /mol_type="genomic DNA"
      /strain="Red Jungle Fowl"
      /db_xref="taxon:9031"
      /clone="CH261-191E5"
      /sex="female"
      /cell_line="UCD001, inbred 256"
      /clone_lib="CH261"
      /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
      CH261 Female Chicken library - For library and clone
      ordering information: http://www.chori.org/bacpac"
ORIGIN
  Query Match 66.7%; Score 20; DB 8; Length 1537;
  Best Local Similarity 82.1%; Pred. No. 2.1e+03;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  QY 2 TCGATCCTTGGCCCGCCCTTGAATTC 29
  Db 1198 TGTGCTTCCCTCCCTCCCTTGAATTC 1171

  RESULT 49
  BX308381/c
  LOCUS
  DEFINITION
    BX308381 tcay Oncorhynchus mykiss cDNA clone tcay0020b.o.11 3prim,
    mRNA sequence.
  ACCESSION
    BX308381
  VERSION
    BX308381.2 GI:40224293
  KEYWORDS
    EST.
  SOURCE
    Oncorhynchus mykiss (rainbow trout)
    ORGANISM
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
      Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
      1 (bases 1 to 669)
      Govoroun,M., Guiguen,Y. and Le Gac,F.
      Construction and primary characterization of normalized cDNA
      libraries in rainbow trout, Oncorhynchus mykiss
      Unpublished (2003)
      On Apr 7, 2003 this sequence version replaced gi:29589026.
      Contact: Guiguen Y
      INRA - SCRIBE
      Campus de Beaulieu, RENNES cedex, 35042, France
      Tel.: 02.23.48.50.09
      Fax: 02.23.48.50.20
      Email: Yann.Guiguen@beaulieu.rennes.inra.fr
      Sequence cleaned of vector, adaptor and repetitions. Contact us
      at signeaupport@jouy.inra.fr to obtain the chromatogram of this
      sequence.
      Plate: 0020 Row: 0 Column: 11
      Seq primer: M13P.
      Location/Qualifiers
        1..669
          /organism="Oncorhynchus mykiss"
          /mol_type="mRNA"
          /db_xref="taxon:8022"
          /clone="tcay0020b.o.11"
          /tissue_type="adipose tissue, blood, brain,
          differentiating gonads, gills, interrenal, intestine,
          kidney, liver, muscle, ovary, pituitary, testis"
  TITLE
  JOURNAL
  COMMENT
  Unpublished (2003)
  Contact: Richard K. Wilson
  Genome Sequencing Center
  Washington University School of Medicine
  Email: submissions@wustl.edu
  Insert Length: 182000 Std Error: 0.00
  Seq primer: Sp6 ATTAGGTGACACTATAG
  Class: BAC ends
  High quality sequence start: 332
  High quality sequence stop: 1022.
  Location/Qualifiers
    1..1537
      /organism="Gallus gallus"
      /mol_type="genomic DNA"
      /strain="Red Jungle Fowl"
      /db_xref="taxon:9031"
      /clone="CH261-191E5"
      /sex="female"
      /cell_line="UCD001, inbred 256"
      /clone_lib="CH261"
      /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
      CH261 Female Chicken library - For library and clone
      ordering information: http://www.chori.org/bacpac"
ORIGIN
  Query Match 66.7%; Score 20; DB 8; Length 1537;
  Best Local Similarity 82.1%; Pred. No. 2.1e+03;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  QY 2 TCGATCCTTGGCCCGCCCTTGAATTC 29
  Db 1198 TGTGCTTCCCTCCCTCCCTTGAATTC 1171

```

/dev_stage="from embryos to adults"

/lab_host="DH10B"

/clone_lib="tcay"

/note="Vector: pTT3D-pac; Rainbow trout multi-tissues -
normalized + 1 subtraction (tcay); Clone distribution :
AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, France, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

ORIGIN

Query Match 66.0%; Score 19.8; DB 5; Length 669;
Best Local Similarity 91.3%; Pred. No. 2.3e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTG 23
|||||
Db 106 CTGGATCCTTGGCCCGCCCTTG 84

RESULT 50

CB509377 714 bp mRNA linear EST 16-MAY-2003
ssalnh503017 whole Salmo salar cDNA, mRNA sequence.

DEFINITION CB509377

ACCESSION CB509377

VERSION CB509377.1 GI:29320603

KEYWORDS EST.

SOURCE Salmo salar (Atlantic salmon)

ORGANISM Salmo salar

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

REFERENCE 1 (bases 1 to 714)

GRASP Consortium, Davidson, W.S., Koop, B.F. and

http://web.uvic.ca/cbr/grasp.

A survey of Salmo salar transcripts from high complexity cDNA

libraries

Unpublished (2002)

Contact: Koop BF

Centre for Biomedical Research

University of Victoria

PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada

Tel: 250 472 4067

Fax: 250 472 4075

Email: bkoop@uvic.ca

Genome Sciences Centre, BC Cancer Agency cDNA preparation,
sequencing and bioinformatics: Y Butterfield, R Kirkpatrick, J
Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prahbu, D
Smallus, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and
M Warra.

POLYA=yes.

FEATURES

source

1..714
/location="Qualifiers
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="whole"
/note="Vector: pBlueScriptISK+; Library Creator: Matthew
L Rise; Atlantic salmon tissue contributors: Carlo Biagi,
Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
B.C.), Rachel Roper (University of Victoria)"

ORIGIN

Query Match 66.0%; Score 19.8; DB 6; Length 714;
Best Local Similarity 91.3%; Pred. No. 2.3e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTG 23
|||||
Db 455 CTGGATCCTTGGCCCGCCCTTG 477

Search completed: April 25, 2005, 15:03:36
Job time : 2608.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:14:29 ; Search time 90 Seconds
(without alignments)
545.426 Million cell updates/sec

Title: US-10-010-476-10
Perfect score: 30
Sequence: 1 CTGGATCTTGGCCCGCCCTTGAATGCC 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A-COMB.seq:
2: /cgn2_6/prodata/1/ina/5B-COMB.seq:
3: /cgn2_6/prodata/1/ina/6A-COMB.seq:
4: /cgn2_6/prodata/1/ina/6B-COMB.seq:
5: /cgn2_6/prodata/1/ina/PCTUS-COMB.seq:
6: /cgn2_6/prodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.6	68.7	122772	4	US-09-949-016-14132
C 2	19.6	65.3	5712	4	US-09-949-016-2417
C 3	19.6	65.3	40435	4	US-09-949-016-16504
C 4	19.6	65.3	57054	4	US-09-949-016-14159
C 5	19.4	64.7	33	3	US-08-840-316-62
C 6	19.4	64.7	33	3	US-08-809-523-62
C 7	19.4	64.7	33	3	US-08-471-971-62
C 8	19.4	64.7	33	3	US-09-402-776-62
C 9	19.4	64.7	33	4	US-08-470-246-62
C 10	19.4	64.7	33	4	US-08-316-765-62
C 11	19.4	64.7	33	4	US-09-724-475-62
C 12	19.4	64.7	33	5	PCT-US93-08849A-62
C 13	19.4	64.7	325034	3	PCT-US93-08849A-62
C 14	19.4	64.7	389504	4	US-09-949-016-14957
C 15	18.8	62.7	37711	4	US-09-949-016-11774
C 16	18.8	62.7	37712	4	US-09-949-016-12832
C 17	18.8	62.7	37712	4	US-09-949-016-16704
C 18	18.8	62.7	50341	1	US-08-247-901C-1
C 19	18.8	62.7	50341	2	US-09-075-904-1
C 20	18.8	62.7	52297	3	US-09-426-436-1
C 21	18.8	62.7	52297	3	US-08-705-557-1
C 22	18.6	62.0	7517	4	US-09-949-016-15603
C 23	18.4	61.3	11105	4	US-09-949-016-13218
C 24	18.4	61.3	12163	4	US-09-949-016-16030
C 25	18.4	61.3	12163	4	US-09-949-016-12202
C 26	18.4	61.3	13426	4	US-09-949-016-13683
C 27	18.4	61.3	71251	4	US-09-949-016-15332

C 28	18.4	61.3	178883	4	US-09-949-016-12733	Sequence 12733, A
C 29	18.4	61.3	178884	4	US-09-949-016-13039	Sequence 13039, A
C 30	18.2	60.7	114426	4	US-09-949-016-15078	Sequence 15078, A
C 31	18	60.0	601	4	US-09-949-016-142664	Sequence 142664, A
C 32	18	60.0	1177	2	US-08-890-542A-1	Sequence 1, Appl1
C 33	18	60.0	1177	2	US-08-890-542A-3	Sequence 3, Appl1
C 34	18	60.0	43307	4	US-09-949-016-15774	Sequence 15774, A
C 35	17.8	59.3	517	4	US-09-621-976-1833	Sequence 1833, Ap
C 36	17.8	59.3	601	4	US-09-949-016-25546	Sequence 25546, A
C 37	17.8	59.3	601	4	US-09-949-016-77051	Sequence 77051, A
C 38	17.8	59.3	601	4	US-09-949-016-131200	Sequence 131200, A
C 39	17.8	59.3	601	4	US-09-949-016-203250	Sequence 203250, A
C 40	17.8	59.3	601	4	US-09-949-016-203251	Sequence 203251, A
C 41	17.8	59.3	601	4	US-09-949-016-203252	Sequence 203252, A
C 42	17.8	59.3	20565	4	US-09-949-016-12125	Sequence 12125, A
C 43	17.8	59.3	20566	4	US-09-949-016-13990	Sequence 13990, A
C 44	17.8	59.3	23928	4	US-09-949-016-12392	Sequence 12392, A
C 45	17.8	59.3	23928	4	US-09-949-016-16315	Sequence 16315, A
C 46	17.8	59.3	26115	4	US-09-949-016-15959	Sequence 15959, A
C 47	17.8	59.3	27589	4	US-09-949-016-17477	Sequence 17477, A
C 48	17.8	59.3	37254	4	US-09-949-016-15973	Sequence 15973, A
C 49	17.8	59.3	72549	4	US-09-949-016-16477	Sequence 16477, A
C 50	17.8	59.3	123463	4	US-09-949-016-17078	Sequence 17078, A
C 51	17.8	59.3	131332	4	US-09-949-016-15535	Sequence 15535, A
C 52	17.8	59.3	199945	4	US-09-949-016-15436	Sequence 15436, A
C 53	17.6	58.7	669	4	US-09-902-540-3536	Sequence 3536, Ap
C 54	17.6	58.7	846	4	US-09-902-540-7470	Sequence 7470, Ap
C 55	17.6	58.7	1491	4	US-09-902-540-4819	Sequence 4819, Ap
C 56	17.6	58.7	4483	4	US-09-799-451-390	Sequence 390, App
C 57	17.6	58.7	5787	4	US-09-902-540-722	Sequence 722, App
C 58	17.6	58.7	8461	4	US-09-949-016-13428	Sequence 13428, A
C 59	17.6	58.7	12151	4	US-09-949-016-16588	Sequence 16588, A
C 60	17.6	58.7	12708	4	US-09-949-016-13216	Sequence 13216, A
C 61	17.6	58.7	18538	4	US-09-902-540-1169	Sequence 1169, Ap
C 62	17.6	58.7	26659	4	US-09-902-540-1237	Sequence 1237, Ap
C 63	17.6	58.7	110585	4	US-09-949-016-13427	Sequence 13427, A
C 64	17.6	58.7	110585	4	US-09-949-016-13427	Sequence 13427, A
C 65	17.6	58.7	373182	4	US-09-949-016-17371	Sequence 17371, A
C 66	17.6	58.7	373694	4	US-09-949-016-12062	Sequence 12062, A
C 67	17.6	58.7	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
C 68	17.6	58.7	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
C 69	17.4	58.0	715	3	US-08-998-416-709	Sequence 709, App
C 70	17.4	58.0	783	4	US-09-252-991A-16173	Sequence 16173, A
C 71	17.4	58.0	851	4	US-09-270-767-29167	Sequence 29167, A
C 72	17.4	58.0	1024	4	US-09-328-475C-29	Sequence 29, Appl
C 73	17.4	58.0	1217	4	US-09-270-767-13239	Sequence 13239, A
C 74	17.4	58.0	1737	4	US-09-270-767-17372	Sequence 17372, A
C 75	17.4	58.0	2091	3	US-08-899-437-22	Sequence 22, Appl
C 76	17.4	58.0	2091	3	US-09-126-121-22	Sequence 5, Appl1
C 77	17.4	58.0	2502	3	US-08-899-437-5	Sequence 5, Appl1
C 78	17.4	58.0	2502	3	US-09-126-121-5	Sequence 1, Appl1
C 79	17.4	58.0	5917	3	US-08-692-922-1	Sequence 1, Appl1
C 80	17.4	58.0	26438	4	US-09-949-016-17351	Sequence 17351, A
C 81	17.4	58.0	32023	4	US-09-949-016-15430	Sequence 15430, A
C 82	17.4	58.0	40655	4	US-09-949-016-12032	Sequence 12032, A
C 83	17.4	58.0	46555	4	US-09-949-016-15919	Sequence 15919, A
C 84	17.4	58.0	63183	4	US-09-949-016-13047	Sequence 13047, A
C 85	17.4	58.0	63183	4	US-09-949-016-13048	Sequence 13048, A
C 86	17.4	58.0	64291	4	US-09-949-016-16278	Sequence 16278, A
C 87	17.4	58.0	117410	4	US-09-949-016-12262	Sequence 12262, A
C 88	17.4	58.0	141248	4	US-09-949-016-12241	Sequence 12241, A
C 89	17.4	58.0	156324	4	US-09-949-016-13749	Sequence 13749, A
C 90	17.2	57.3	601	4	US-09-949-016-37818	Sequence 37818, A
C 91	17.2	57.3	601	4	US-09-949-016-52216	Sequence 52216, A
C 92	17.2	57.3	601	4	US-09-949-016-85095	Sequence 85095, A
C 93	17.2	57.3	601	4	US-09-949-016-85096	Sequence 85096, A
C 94	17.2	57.3	601	4	US-09-949-016-85246	Sequence 85246, A
C 95	17.2	57.3	601	4	US-09-949-016-85247	Sequence 85247, A
C 96	17.2	57.3	601	4	US-09-949-016-177932	Sequence 177932, A
C 97	17.2	57.3	601	4	US-09-949-016-177939	Sequence 177939, A
C 98	17.2	57.3	883	4	US-09-902-540-100	Sequence 100, App
C 99	17.2	57.3	1050	4	US-09-614-221A-349	Sequence 349, App
C 100	17.2	57.3	1319	3	US-09-009-816-20	Sequence 20, Appl

c 101	17.2	57.3	1856	4	US-09-902-540-6531	Sequence 6531, Ap	c 174	16.8	55.0	330	4	US-09-902-540-8688	Sequence 8688, Ap
c 102	17.2	57.3	5703	3	US-09-280-590A-36	Sequence 36, Appl	175	16.8	56.0	520	3	US-09-247-155-135	Sequence 135, App
c 103	17.2	57.3	5703	4	US-09-892-398-36	Sequence 36, Appl	176	16.8	56.0	556	3	US-09-370-838-242	Sequence 242, App
c 104	17.2	57.3	8879	4	US-09-949-016-11827	Sequence 11827, A	177	16.8	56.0	556	4	US-09-854-133-242	Sequence 242, App
c 105	17.2	57.3	8880	4	US-09-949-016-16003	Sequence 16003, A	178	16.8	56.0	601	4	US-09-949-016-27033	Sequence 27033, A
c 106	17.2	57.3	10182	4	US-09-902-540-10049	Sequence 1049, Ap	179	16.8	56.0	601	4	US-09-949-016-30367	Sequence 30367, A
c 107	17.2	57.3	11601	2	US-08-222-617A-3	Sequence 3, Appl1	c 180	16.8	56.0	601	4	US-09-949-016-48295	Sequence 48295, A
c 108	17.2	57.3	11601	2	US-08-222-617A-24	Sequence 24, Appl1	181	16.8	56.0	601	4	US-09-949-016-48295	Sequence 48295, A
c 109	17.2	57.3	12249	4	US-09-949-016-16839	Sequence 16839, A	182	16.8	56.0	601	4	US-09-949-016-57396	Sequence 57396, A
c 110	17.2	57.3	12249	4	US-09-949-016-16840	Sequence 16840, A	183	16.8	56.0	601	4	US-09-949-016-65408	Sequence 65408, A
c 111	17.2	57.3	12928	4	US-09-949-016-12772	Sequence 12772, A	184	16.8	56.0	601	4	US-09-949-016-69656	Sequence 69656, A
c 112	17.2	57.3	12929	4	US-09-949-016-16702	Sequence 16702, A	c 185	16.8	56.0	601	4	US-09-949-016-124017	Sequence 124017, A
c 113	17.2	57.3	21521	4	US-09-949-016-14294	Sequence 14294, A	186	16.8	56.0	601	4	US-09-949-016-124018	Sequence 124018, A
c 114	17.2	57.3	34266	4	US-09-949-016-13250	Sequence 13250, A	c 187	16.8	56.0	601	4	US-09-949-016-124019	Sequence 124019, A
c 115	17.2	57.3	36180	4	US-09-949-016-11745	Sequence 11745, A	188	16.8	56.0	601	4	US-09-949-016-137632	Sequence 137632, A
c 116	17.2	57.3	36181	4	US-09-949-016-16163	Sequence 16163, A	c 189	16.8	56.0	712	1	US-08-340-539A-7	Sequence 7, Appl1
c 117	17.2	57.3	36223	4	US-09-949-016-14417	Sequence 14417, A	190	16.8	56.0	712	2	US-08-461-592B-7	Sequence 7, Appl1
c 118	17.2	57.3	36741	3	US-09-301-665-3	Sequence 3, Appl1	c 191	16.8	56.0	732	4	US-09-854-133-429	Sequence 429, App
c 119	17.2	57.3	48536	4	US-09-949-016-11867	Sequence 11867, A	192	16.8	56.0	732	4	US-09-253-991A-9733	Sequence 9733, Ap
c 120	17.2	57.3	48536	4	US-09-949-016-11867	Sequence 11867, A	c 193	16.8	56.0	970	4	US-09-799-451-263	Sequence 263, App
c 121	17.2	57.3	49372	1	US-08-614-770A-1	Sequence 1, Appl1	194	16.8	56.0	1089	4	US-09-252-991A-9606	Sequence 9606, Ap
c 122	17.2	57.3	49378	4	US-09-949-016-13408	Sequence 13408, A	c 194	16.8	56.0	1307	4	US-09-023-655-293	Sequence 293, App
c 123	17.2	57.3	77826	4	US-09-949-016-12608	Sequence 12608, A	c 195	16.8	56.0	1307	4	US-09-252-991A-9561	Sequence 9561, Ap
c 124	17.2	57.3	86127	4	US-09-949-016-12209	Sequence 12209, A	c 196	16.8	56.0	1335	4	US-09-252-991A-9690	Sequence 9690, Ap
c 125	17.2	57.3	86127	4	US-09-949-016-12209	Sequence 12209, A	c 197	16.8	56.0	2790	4	US-09-904-615-30	Sequence 30, Appl1
c 126	17.2	57.3	91933	4	US-09-949-016-12757	Sequence 12757, A	c 198	16.8	56.0	3068	1	US-07-984-044A-3	Sequence 3, Appl1
c 127	17.2	57.3	91933	4	US-09-949-016-12757	Sequence 12757, A	199	16.8	56.0	3068	1	US-08-458-393-3	Sequence 3, Appl1
c 128	17.2	57.3	109159	4	US-09-949-016-14169	Sequence 14169, A	200	16.8	56.0	3068	1	US-08-630-915A-33	Sequence 33, Appl1
c 129	17.2	57.3	109159	4	US-09-949-016-14170	Sequence 14170, A	c 201	16.8	56.0	4091	4	US-09-879-957-33	Sequence 33, Appl1
c 130	17.2	57.3	139150	4	US-09-949-016-17398	Sequence 17398, A	c 202	16.8	56.0	4091	4	US-09-879-957-33	Sequence 33, Appl1
c 131	17.2	57.3	139150	4	US-09-949-016-17398	Sequence 17398, A	c 203	16.8	56.0	5884	4	US-09-949-016-17129	Sequence 17129, A
c 132	17.2	57.3	268449	4	US-09-949-016-17244	Sequence 17244, A	c 204	16.8	56.0	7553	4	US-09-902-540-920	Sequence 920, App
c 133	17.2	57.3	323820	4	US-09-949-016-14139	Sequence 14139, A	c 205	16.8	56.0	7553	4	US-09-902-540-954	Sequence 954, App
c 134	17.2	56.7	204	4	US-09-513-999C-35308	Sequence 35308, A	c 206	16.8	56.0	8967	4	US-09-902-540-954	Sequence 954, App
c 135	17.2	56.7	261	4	US-09-513-999C-32493	Sequence 32493, A	c 207	16.8	56.0	10082	4	US-09-949-016-15569	Sequence 15569, A
c 136	17.2	56.7	563	4	US-09-270-767-1132	Sequence 1132, Ap	c 208	16.8	56.0	11089	4	US-09-949-016-12381	Sequence 12381, A
c 137	17.2	56.7	563	4	US-09-424-978B-25	Sequence 25, Appl	c 209	16.8	56.0	11089	4	US-09-949-016-12381	Sequence 12381, A
c 138	17.2	56.7	643	4	US-09-134-000C-1262	Sequence 1262, Ap	c 210	16.8	56.0	11865	4	US-09-949-016-15297	Sequence 15297, A
c 139	17.2	56.7	678	4	US-08-413-974-3	Sequence 3, Appl1	c 211	16.8	56.0	11865	4	US-09-949-016-15297	Sequence 15297, A
c 140	17.2	56.7	756	3	US-08-434-418-3	Sequence 3, Appl1	c 212	16.8	56.0	11865	4	US-09-949-016-15297	Sequence 15297, A
c 141	17.2	56.7	756	3	US-08-434-288-3	Sequence 3, Appl1	c 213	16.8	56.0	2166	4	US-09-949-016-13755	Sequence 13755, A
c 142	17.2	56.7	756	3	US-08-434-288-3	Sequence 3, Appl1	c 214	16.8	56.0	2166	4	US-09-949-016-13755	Sequence 13755, A
c 143	17.2	56.7	756	3	US-08-434-256-3	Sequence 3, Appl1	c 215	16.8	56.0	36620	4	US-09-952-060-31	Sequence 31, Appl
c 144	17.2	56.7	810	3	US-08-413-974-5	Sequence 5, Appl1	c 216	16.8	56.0	36620	4	US-09-952-060-31	Sequence 31, Appl
c 145	17.2	56.7	810	3	US-08-434-418-5	Sequence 5, Appl1	c 217	16.8	56.0	38479	4	US-09-949-016-16730	Sequence 16730, A
c 146	17.2	56.7	810	3	US-08-434-288-5	Sequence 5, Appl1	c 218	16.8	56.0	38479	4	US-09-952-060-29	Sequence 29, Appl
c 147	17.2	56.7	810	3	US-08-434-288-5	Sequence 5, Appl1	c 219	16.8	56.0	38528	4	US-09-949-016-11790	Sequence 11790, A
c 148	17.2	56.7	810	3	US-08-434-288-5	Sequence 5, Appl1	c 220	16.8	56.0	38528	4	US-09-949-016-15475	Sequence 15475, A
c 149	17.2	56.7	810	3	US-08-434-288-5	Sequence 5, Appl1	c 221	16.8	56.0	45225	4	US-09-949-016-12428	Sequence 12428, A
c 150	17.2	56.7	1123	1	US-07-971-096-3	Sequence 3, Appl1	c 222	16.8	56.0	45225	4	US-09-949-016-12428	Sequence 12428, A
c 151	17.2	56.7	1123	1	US-08-175-096-3	Sequence 3, Appl1	c 223	16.8	56.0	49136	3	US-09-422-869-1	Sequence 1, Appl1
c 152	17.2	56.7	1644	4	US-09-902-540-371	Sequence 371, App	c 224	16.8	56.0	49136	3	US-09-422-869-1	Sequence 1, Appl1
c 153	17.2	56.7	3719	1	US-08-920-810-10	Sequence 10, Appl	c 225	16.8	56.0	57320	4	US-09-949-016-13983	Sequence 13983, A
c 154	17.2	56.7	3719	1	US-08-920-810-10	Sequence 10, Appl	c 226	16.8	56.0	57320	4	US-09-949-016-12234	Sequence 12234, A
c 155	17.2	56.7	3719	1	US-08-362-577C-10	Sequence 10, Appl	c 227	16.8	56.0	91665	4	US-09-949-016-12234	Sequence 12234, A
c 156	17.2	56.7	3719	2	US-08-920-810-10	Sequence 10, Appl	c 228	16.8	56.0	91665	4	US-09-949-016-12234	Sequence 12234, A
c 157	17.2	56.7	7353	4	US-09-949-016-14895	Sequence 14895, A	c 229	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 158	17.2	56.7	9278	1	US-08-243-544-9	Sequence 9, Appl1	c 230	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 159	17.2	56.7	9278	1	US-08-477-407-9	Sequence 9, Appl1	c 231	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 160	17.2	56.7	9278	1	US-08-484-355-9	Sequence 9, Appl1	c 232	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 161	17.2	56.7	9364	4	US-09-949-016-15965	Sequence 15965, A	c 233	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 162	17.2	56.7	17138	3	US-09-813-819-3	Sequence 3, Appl1	c 234	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 163	17.2	56.7	17138	3	US-09-813-819-3	Sequence 3, Appl1	c 235	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 164	17.2	56.7	17138	3	US-09-813-819-3	Sequence 3, Appl1	c 236	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 165	17.2	56.7	18157	4	US-09-949-016-16193	Sequence 16193, A	c 237	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 166	17.2	56.7	36811	4	US-09-949-016-17287	Sequence 17287, A	c 238	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 167	17.2	56.7	45275	4	US-09-949-016-12533	Sequence 12533, A	c 239	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 168	17.2	56.7	80246	3	US-09-078-294-4	Sequence 4, Appl1	c 240	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 169	17.2	56.7	80595	3	US-09-078-294-4	Sequence 4, Appl1	c 241	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 170	17.2	56.7	84296	4	US-09-949-016-17375	Sequence 17375, A	c 242	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 171	17.2	56.7	95122	4	US-09-949-016-17235	Sequence 17235, A	c 243	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 172	17.2	56.7	153866	4	US-09-949-016-16919	Sequence 16919, A	c 244	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A
c 173	17.2	56.7	276237	4	US-09-949-016-17504	Sequence 17504, A	c 245	16.8	56.0	109038	4	US-09-949-016-12199	Sequence 12199, A

C 247	16.6	55.3	601	4	US-09-949-016-81525	Sequence 81525, A	C 320	16.4	54.7	76247	4	US-09-949-016-16136	Sequence 16136, A
C 248	16.6	55.3	601	4	US-09-949-016-120050	Sequence 120050, A	C 321	16.4	54.7	85369	4	US-09-949-016-12171	Sequence 12171, A
C 249	16.6	55.3	601	4	US-09-949-016-120100	Sequence 120100, A	C 322	16.4	54.7	175236	4	US-09-949-016-13533	Sequence 13533, A
C 250	16.6	55.3	601	4	US-09-949-016-120150	Sequence 120150, A	C 323	16.4	54.7	312470	4	US-09-949-016-14043	Sequence 14043, A
C 251	16.6	55.3	601	4	US-09-949-016-127859	Sequence 127859, A	C 324	16.4	54.7	325791	4	US-09-768-185A-1	Sequence 1, Appli
C 252	16.6	55.3	601	4	US-09-949-016-128196	Sequence 128196, A	C 325	16.4	54.7	390416	4	US-09-949-016-15923	Sequence 16923, A
C 253	16.6	55.3	685	4	US-09-620-312D-806	Sequence 806, App	C 326	16.4	54.7	536165	4	US-09-214-808-1	Sequence 1, Appli
C 254	16.6	55.3	727	4	US-09-620-312D-805	Sequence 805, App	C 327	16.4	54.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 255	16.6	55.3	729	4	US-09-583-110-324	Sequence 324, App	C 328	16.4	54.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 256	16.6	55.3	738	4	US-09-107-433-2224	Sequence 2224, Ap	C 329	16.2	54.0	30	3	US-08-840-316-65	Sequence 102, App
C 257	16.6	55.3	740	4	US-09-370-767-27636	Sequence 27636, A	C 330	16.2	54.0	30	3	US-08-840-316-102	Sequence 65, Appl
C 258	16.6	55.3	858	4	US-09-302-540-6737	Sequence 6737, Ap	C 331	16.2	54.0	30	3	US-08-809-523-65	Sequence 65, Appl
C 259	16.6	55.3	1155	4	US-09-491-577-15	Sequence 15, Appl	C 332	16.2	54.0	30	3	US-08-809-523-102	Sequence 102, App
C 260	16.6	55.3	1354	4	US-09-491-577-106	Sequence 106, App	C 333	16.2	54.0	30	3	US-08-471-971-65	Sequence 65, Appl
C 261	16.6	55.3	1340	4	US-09-370-767-11959	Sequence 11959, A	C 334	16.2	54.0	30	3	US-08-471-971-102	Sequence 102, App
C 262	16.6	55.3	4009	3	US-09-302-540-541	Sequence 541, App	C 335	16.2	54.0	30	3	US-09-402-776-65	Sequence 65, Appl
C 263	16.6	55.3	4739	3	US-08-677-970-1	Sequence 1, Appli	C 336	16.2	54.0	30	3	US-09-402-776-102	Sequence 102, App
C 264	16.6	55.3	5481	4	US-09-302-540-7351	Sequence 7351, Ap	C 337	16.2	54.0	30	4	US-08-470-246-65	Sequence 65, Appl
C 265	16.6	55.3	6273	3	US-08-961-527-21	Sequence 21, Appl	C 338	16.2	54.0	30	4	US-08-470-246-102	Sequence 102, App
C 266	16.6	55.3	6402	4	US-09-902-540-702	Sequence 702, App	C 339	16.2	54.0	30	4	US-08-316-765-65	Sequence 65, Appl
C 267	16.6	55.3	7261	4	US-09-949-016-14780	Sequence 14780, A	C 340	16.2	54.0	30	4	US-08-316-765-102	Sequence 102, App
C 268	16.6	55.3	19319	4	US-09-949-016-14107	Sequence 14107, A	C 341	16.2	54.0	30	4	US-09-724-475-65	Sequence 65, Appl
C 269	16.6	55.3	28229	4	US-09-949-016-12368	Sequence 12368, A	C 342	16.2	54.0	30	4	US-09-724-475-102	Sequence 102, App
C 270	16.6	55.3	44393	4	US-09-949-016-14944	Sequence 14944, A	C 343	16.2	54.0	30	5	PCT-US93-08849A-65	Sequence 65, Appl
C 271	16.6	55.3	44393	4	US-09-949-016-14945	Sequence 14945, A	C 344	16.2	54.0	30	5	PCT-US93-08849-65	Sequence 65, Appl
C 272	16.6	55.3	44393	4	US-09-949-016-14946	Sequence 14946, A	C 345	16.2	54.0	279	3	US-09-643-597-200	Sequence 200, App
C 273	16.6	55.3	44393	4	US-09-949-016-16911	Sequence 16911, A	C 346	16.2	54.0	279	4	US-09-480-884A-200	Sequence 200, App
C 274	16.6	55.3	47476	4	US-09-949-016-12179	Sequence 12179, A	C 347	16.2	54.0	279	4	US-09-543-615A-200	Sequence 200, App
C 275	16.6	55.3	47476	4	US-09-949-016-14472	Sequence 14472, A	C 348	16.2	54.0	279	4	US-09-606-421B-200	Sequence 200, App
C 276	16.6	55.3	50850	4	US-09-949-016-15083	Sequence 15083, A	C 349	16.2	54.0	279	4	US-09-466-396A-200	Sequence 200, App
C 277	16.6	55.3	50850	4	US-09-949-016-15084	Sequence 15084, A	C 350	16.2	54.0	279	4	US-09-476-496A-200	Sequence 200, App
C 278	16.6	55.3	50850	4	US-09-949-016-15085	Sequence 15085, A	C 351	16.2	54.0	279	4	US-09-630-940B-200	Sequence 200, App
C 279	16.6	55.3	90900	4	US-09-949-016-13034	Sequence 13034, A	C 352	16.2	54.0	329	4	US-09-513-993C-16254	Sequence 16254, A
C 280	16.6	55.3	176006	4	US-09-949-016-16804	Sequence 16804, A	C 353	16.2	54.0	342	4	US-09-107-532A-742	Sequence 742, App
C 281	16.6	55.3	253375	4	US-09-949-016-12849	Sequence 12849, A	C 354	16.2	54.0	364	4	US-09-270-767-9531	Sequence 9531, Ap
C 282	16.6	55.3	304533	4	US-09-949-016-15371	Sequence 15371, A	C 355	16.2	54.0	364	4	US-09-270-767-24813	Sequence 24813, A
C 283	16.6	55.3	304533	4	US-09-949-016-15372	Sequence 15372, A	C 356	16.2	54.0	404	4	US-09-621-976-15291	Sequence 15291, A
C 284	16.6	55.3	784019	4	US-09-949-016-14033	Sequence 14033, A	C 357	16.2	54.0	441	1	US-08-217-918-3	Sequence 3, Appli
C 285	16.6	55.3	828152	4	US-09-949-016-12777	Sequence 12777, A	C 358	16.2	54.0	502	4	US-09-621-976-2480	Sequence 2480, Ap
C 286	16.4	54.7	135	4	US-09-495-050A-110	Sequence 110, App	C 359	16.2	54.0	506	4	US-09-621-976-614	Sequence 614, App
C 287	16.4	54.7	215	4	US-09-270-767-6673	Sequence 6673, Ap	C 360	16.2	54.0	525	4	US-09-854-133-644	Sequence 644, App
C 288	16.4	54.7	535	4	US-09-270-767-6673	Sequence 6673, Ap	C 361	16.2	54.0	601	4	US-09-949-016-12924	Sequence 12924, A
C 289	16.4	54.7	535	4	US-09-270-767-21955	Sequence 21955, A	C 362	16.2	54.0	601	4	US-09-949-016-12925	Sequence 12925, A
C 290	16.4	54.7	601	4	US-09-949-016-60569	Sequence 60569, A	C 363	16.2	54.0	601	4	US-09-949-016-26417	Sequence 26417, A
C 291	16.4	54.7	601	4	US-09-949-016-79935	Sequence 79935, A	C 364	16.2	54.0	601	4	US-09-949-016-29485	Sequence 29485, A
C 292	16.4	54.7	601	4	US-09-949-016-79936	Sequence 79936, A	C 365	16.2	54.0	601	4	US-09-949-016-30439	Sequence 30439, A
C 293	16.4	54.7	601	4	US-09-949-016-130348	Sequence 130348, A	C 366	16.2	54.0	601	4	US-09-949-016-30440	Sequence 30440, A
C 294	16.4	54.7	601	4	US-09-949-016-136798	Sequence 136798, A	C 367	16.2	54.0	601	4	US-09-949-016-30441	Sequence 30441, A
C 295	16.4	54.7	735	3	US-09-484-970B-71	Sequence 71, Appl	C 368	16.2	54.0	601	4	US-09-949-016-33690	Sequence 33690, A
C 296	16.4	54.7	1001	3	US-09-641-638-175	Sequence 175, App	C 369	16.2	54.0	601	4	US-09-949-016-33691	Sequence 33691, A
C 297	16.4	54.7	1001	4	US-10-170-097-175	Sequence 175, App	C 370	16.2	54.0	601	4	US-09-949-016-33699	Sequence 33699, A
C 298	16.4	54.7	1140	4	US-09-491-577-27	Sequence 27, Appl	C 371	16.2	54.0	601	4	US-09-949-016-38659	Sequence 38659, A
C 299	16.4	54.7	1140	4	US-09-491-577-112	Sequence 112, App	C 372	16.2	54.0	601	4	US-09-949-016-38660	Sequence 38660, A
C 300	16.4	54.7	1843	4	US-09-023-655-753	Sequence 753, App	C 373	16.2	54.0	601	4	US-09-949-016-38661	Sequence 38661, A
C 301	16.4	54.7	2184	4	US-09-902-540-5116	Sequence 5116, Ap	C 374	16.2	54.0	601	4	US-09-949-016-46497	Sequence 46497, A
C 302	16.4	54.7	2527	4	US-09-949-016-3838	Sequence 3838, Ap	C 375	16.2	54.0	601	4	US-09-949-016-64698	Sequence 64698, A
C 303	16.4	54.7	2625	3	US-09-369-364A-14	Sequence 14, Appl	C 376	16.2	54.0	601	4	US-09-949-016-70778	Sequence 70778, A
C 304	16.4	54.7	3024	1	US-08-149-100-1	Sequence 1, Appli	C 377	16.2	54.0	601	4	US-09-949-016-80284	Sequence 80284, A
C 305	16.4	54.7	3729	4	US-09-107-532A-1587	Sequence 1587, Ap	C 378	16.2	54.0	601	4	US-09-949-016-88195	Sequence 88195, A
C 306	16.4	54.7	9700	4	US-09-698-293-9	Sequence 9, Appli	C 379	16.2	54.0	601	4	US-09-949-016-90154	Sequence 90154, A
C 307	16.4	54.7	9865	4	US-09-698-293-2	Sequence 2, Appli	C 380	16.2	54.0	601	4	US-09-949-016-90155	Sequence 90155, A
C 308	16.4	54.7	18112	4	US-09-949-016-13648	Sequence 13648, A	C 381	16.2	54.0	601	4	US-09-949-016-90156	Sequence 90156, A
C 309	16.4	54.7	28058	4	US-09-902-540-1352	Sequence 1252, Ap	C 382	16.2	54.0	601	4	US-09-949-016-90180	Sequence 90180, A
C 310	16.4	54.7	28438	4	US-09-820-790B-3	Sequence 3, Appli	C 383	16.2	54.0	601	4	US-09-949-016-90181	Sequence 90181, A
C 311	16.4	54.7	31623	4	US-09-949-016-15945	Sequence 15945, A	C 384	16.2	54.0	601	4	US-09-949-016-90182	Sequence 90182, A
C 312	16.4	54.7	36620	4	US-09-952-060-31	Sequence 31, Appl	C 385	16.2	54.0	601	4	US-09-949-016-94351	Sequence 94351, A
C 313	16.4	54.7	37474	4	US-09-952-060-26	Sequence 26, Appl	C 386	16.2	54.0	601	4	US-09-949-016-131529	Sequence 131529, A
C 314	16.4	54.7	38155	3	US-09-453-702B-79	Sequence 79, Appl	C 387	16.2	54.0	601	4	US-09-949-016-131530	Sequence 131530, A
C 315	16.4	54.7	38519	4	US-09-952-060-29	Sequence 29, Appl	C 388	16.2	54.0	601	4	US-09-949-016-142254	Sequence 142254, A
C 316	16.4	54.7	54420	4	US-09-949-016-15422	Sequence 15422, A	C 389	16.2	54.0	601	4	US-09-949-016-142255	Sequence 142255, A
C 317	16.4	54.7	67002	4	US-09-949-016-16803	Sequence 16803, A	C 390	16.2	54.0	601	4	US-09-949-016-156429	Sequence 156429, A
C 318	16.4	54.7	71574	4	US-09-949-016-15580	Sequence 15580, A	C 391	16.2	54.0	601	4	US-09-949-016-156430	Sequence 156430, A
C 319	16.4	54.7	74644	4	US-09-949-016-17556	Sequence 17556, A	C 392	16.2	54.0	601	4	US-09-949-016-167221	Sequence 167221, A

C 393	US-09-632-945-1	4	648	54.0	16.2	Sequence 1, Appli	466	16.2	54.0	107140	4	US-09-949-016-14934	Sequence 14834, A
C 394	US-09-328-475C-97	4	1024	54.0	16.2	Sequence 97, Appl	C 467	16.2	54.0	113283	4	US-09-949-016-16976	Sequence 16976, A
C 395	US-09-328-475C-98	4	1024	54.0	16.2	Sequence 98, Appl	C 468	16.2	54.0	113283	4	US-09-949-016-16977	Sequence 16977, A
C 396	US-09-902-540-3138	4	1046	54.0	16.2	Sequence 3138, Ap	C 469	16.2	54.0	115643	4	US-09-949-016-12298	Sequence 12298, A
C 397	US-09-799-451-651	4	1085	54.0	16.2	Sequence 651, App	C 470	16.2	54.0	115649	4	US-09-949-016-12537	Sequence 12537, A
C 398	US-09-799-451-694	4	1118	54.0	16.2	Sequence 694, App	C 471	16.2	54.0	124700	4	US-09-949-016-11817	Sequence 11817, A
C 399	US-09-949-016-4713	4	1207	54.0	16.2	Sequence 4713, Ap	C 472	16.2	54.0	124701	4	US-09-949-016-15439	Sequence 15439, A
C 400	US-09-949-016-594	4	1223	54.0	16.2	Sequence 594, App	C 473	16.2	54.0	137226	4	US-09-949-016-13763	Sequence 13763, A
C 401	US-09-134-000C-2669	4	1368	54.0	16.2	Sequence 2669, Ap	C 474	16.2	54.0	137226	4	US-09-949-016-13763	Sequence 13763, A
C 402	US-09-976-594-50	4	1577	54.0	16.2	Sequence 50, Appl	C 475	16.2	54.0	139936	4	US-09-949-016-11782	Sequence 11782, A
C 403	US-09-023-555-878	4	1587	54.0	16.2	Sequence 878, App	C 476	16.2	54.0	144322	4	US-09-949-016-13280	Sequence 13280, A
C 404	US-09-673-300-9	4	1669	54.0	16.2	Sequence 9, Appli	C 477	16.2	54.0	147382	4	US-09-949-016-15316	Sequence 15316, A
C 405	US-09-443-041A-21	3	1695	54.0	16.2	Sequence 21, Appl	C 478	16.2	54.0	147382	4	US-09-949-016-14624	Sequence 14624, A
C 406	US-09-252-991A-490	4	1758	54.0	16.2	Sequence 490, App	C 479	16.2	54.0	171701	4	US-09-949-016-12276	Sequence 12276, A
C 407	US-09-252-991A-568	4	1794	54.0	16.2	Sequence 568, App	C 480	16.2	54.0	199945	4	US-09-949-016-15835	Sequence 15835, A
C 408	US-09-620-312D-313	4	1825	54.0	16.2	Sequence 313, App	C 481	16.2	54.0	203475	4	US-09-949-016-15436	Sequence 15436, A
C 409	US-09-799-451-290	4	2126	54.0	16.2	Sequence 290, App	C 482	16.2	54.0	203475	4	US-09-949-016-14517	Sequence 14517, A
C 410	US-09-277-262-7	3	2240	54.0	16.2	Sequence 7, Appli	C 483	16.2	54.0	203475	4	US-09-949-016-14518	Sequence 14518, A
C 411	US-09-976-594-445	4	2918	54.0	16.2	Sequence 445, App	C 484	16.2	54.0	203475	4	US-09-949-016-14519	Sequence 14519, A
C 412	US-09-184-418C-65	4	3009	54.0	16.2	Sequence 65, Appl	C 485	16.2	54.0	203475	4	US-09-949-016-17226	Sequence 17226, A
C 413	US-07-678-408A-1	1	3944	54.0	16.2	Sequence 1, Appli	C 486	16.2	54.0	203475	4	US-09-949-016-17227	Sequence 17227, A
C 414	US-09-526-193A-16	4	4768	54.0	16.2	Sequence 16, Appl	C 487	16.2	54.0	203475	4	US-09-949-016-17228	Sequence 17228, A
C 415	US-08-973-544-1	3	5581	54.0	16.2	Sequence 1, Appli	C 488	16.2	54.0	203475	4	US-09-949-016-17229	Sequence 17229, A
C 416	US-09-976-594-183	4	5833	54.0	16.2	Sequence 183, App	C 489	16.2	54.0	232024	4	US-09-949-016-13477	Sequence 13477, A
C 417	US-09-949-016-4205	4	6882	54.0	16.2	Sequence 4205, Ap	C 490	16.2	54.0	232452	4	US-09-949-016-13675	Sequence 13675, A
C 418	US-09-949-016-12336	4	8246	54.0	16.2	Sequence 11923, A	C 491	16.2	54.0	254405	4	US-09-949-016-14381	Sequence 14381, A
C 419	US-09-949-016-16648	4	8246	54.0	16.2	Sequence 16648, A	C 492	16.2	54.0	340380	4	US-09-949-016-14179	Sequence 14179, A
C 420	US-09-184-418C-7	4	9050	54.0	16.2	Sequence 7, Appli	C 493	16.2	54.0	421491	4	US-09-949-016-12805	Sequence 12805, A
C 421	US-09-949-016-14299	4	9391	54.0	16.2	Sequence 14299, A	C 494	16.2	54.0	421494	4	US-09-949-016-14060	Sequence 14060, A
C 422	US-09-949-016-13512	4	13445	54.0	16.2	Sequence 13512, A	C 495	16.2	54.0	536165	4	US-09-214-808-1	Sequence 1, Appli
C 423	US-09-949-016-16376	4	13634	54.0	16.2	Sequence 13676, A	C 496	16	53.3	187	3	US-09-106-638-2	Sequence 2, Appli
C 424	US-09-902-540-1137	4	15132	54.0	16.2	Sequence 1137, Ap	C 497	16	53.3	187	4	US-09-513-999C-27535	Sequence 27535, A
C 425	US-09-949-016-17225	4	15543	54.0	16.2	Sequence 17225, A	C 498	16	53.3	210	4	US-09-489-039A-6085	Sequence 6085, Ap
C 426	US-09-949-016-17221	4	15615	54.0	16.2	Sequence 17221, A	C 499	16	53.3	210	4	US-09-248-796A-7453	Sequence 7453, Ap
C 427	US-09-631-320-3	4	17000	54.0	16.2	Sequence 3, Appli	C 500	16	53.3	228	4	US-09-248-796A-13200	Sequence 13200, A
C 428	US-09-902-540-1143	4	19598	54.0	16.2	Sequence 1143, Ap							
C 429	US-09-949-016-12336	4	19993	54.0	16.2	Sequence 12336, A							
C 430	US-09-949-016-16455	4	20137	54.0	16.2	Sequence 16455, A							
C 431	US-09-262-773-206	3	20137	54.0	16.2	Sequence 206, App							
C 432	US-09-262-773-9	3	20138	54.0	16.2	Sequence 9, Appli							
C 433	US-09-262-773-210	3	20138	54.0	16.2	Sequence 10, App							
C 434	US-09-949-016-13207	4	25762	54.0	16.2	Sequence 13207, A							
C 435	US-09-949-016-16263	4	29291	54.0	16.2	Sequence 16263, A							
C 436	US-09-949-016-14213	4	30456	54.0	16.2	Sequence 14213, A							
C 437	US-09-949-016-14049	4	30528	54.0	16.2	Sequence 14049, A							
C 438	US-09-949-016-14449	4	34088	54.0	16.2	Sequence 14449, A							
C 439	US-09-949-016-12384	4	39154	54.0	16.2	Sequence 12384, A							
C 440	US-09-949-016-12801	4	39154	54.0	16.2	Sequence 12801, A							
C 441	US-09-949-016-14326	4	39443	54.0	16.2	Sequence 14326, A							
C 442	US-09-949-016-14327	4	39443	54.0	16.2	Sequence 14327, A							
C 443	US-09-949-016-11932	4	41522	54.0	16.2	Sequence 11932, A							
C 444	US-09-949-016-15764	4	41523	54.0	16.2	Sequence 15764, A							
C 445	US-09-949-016-16296	4	49301	54.0	16.2	Sequence 16296, A							
C 446	US-09-949-016-12547	4	49389	54.0	16.2	Sequence 12547, A							
C 447	US-09-949-016-13630	4	49389	54.0	16.2	Sequence 13630, A							
C 448	US-09-949-016-13184	4	51928	54.0	16.2	Sequence 13184, A							
C 449	US-09-949-016-12130	4	52789	54.0	16.2	Sequence 12130, A							
C 450	US-09-949-016-16641	4	52790	54.0	16.2	Sequence 16641, A							
C 451	US-09-801-861-3	4	53332	54.0	16.2	Sequence 3, Appli							
C 452	US-10-224-562-3	4	53332	54.0	16.2	Sequence 3, Appli							
C 453	US-09-949-016-16286	4	53562	54.0	16.2	Sequence 16286, A							
C 454	US-09-800-960-3	3	62804	54.0	16.2	Sequence 3, Appli							
C 455	US-10-096-960-3	3	62804	54.0	16.2	Sequence 3, Appli							
C 456	US-09-851-896-3	3	70000	54.0	16.2	Sequence 3, Appli							
C 457	US-09-949-016-12126	4	71645	54.0	16.2	Sequence 12126, A							
C 458	US-09-949-016-17258	4	71651	54.0	16.2	Sequence 17258, A							
C 459	US-09-685-853A-3	4	74962	54.0	16.2	Sequence 3, Appli							
C 460	US-09-949-016-16819	4	76399	54.0	16.2	Sequence 16819, A							
C 461	US-09-949-016-12362	4	85912	54.0	16.2	Sequence 12362, A							
C 462	US-09-949-016-16109	4	85913	54.0	16.2	Sequence 16109, A							
C 463	US-09-949-016-17163	4	92074	54.0	16.2	Sequence 17163, A							
C 464	US-09-949-016-15096	4	93778	54.0	16.2	Sequence 15096, A							
C 465	US-09-949-016-13834	4	106924	54.0	16.2	Sequence 13834, A							

ALIGNMENTS

RESULT 1

US-09-949-016-14132/c

Sequence 14132, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14132

LENGTH: 122772

TYPE: DNA

ORGANISM: Human

US-09-949-016-14132

Query Match

Best Local Similarity

Matches

23;

Conservative

0;

Mismatches

4;

Indels

0;

Gaps

0;

QY

1

CTGGATCCTTGCCCGCCCTTGATT

27

DB

48574

CTGGAGCCGAGCCCGCCCTTGATT

48548

SEQUENCE CHARACTERISTICS:
 LENGTH: 33 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-840-316-62

Query Match 64.7%; Score 19.4; DB 3; Length 33;
 Best Local Similarity 79.3%; Pred. No. 46;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
 |||||
 Db 32 CTGGTCTTTATGCGCCCTCGAATTC 4

RESULT 6

US-08-809-523-62/c
 ; Sequence 62, Application US/08809523
 ; Patent No. 6207416

GENERAL INFORMATION:
 APPLICANT: Tsarev, Sergei. A., Emerson,
 APPLICANT: Suzanne U., Purcell, Robert H.
 TITLE OF INVENTION: Recombinant Proteins Of
 TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
 TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 NUMBER OF SEQUENCES: 107
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/809,523
 FILING DATE: 28-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/13102
 FILING DATE: 03-OCT-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US08/316,765
 FILING DATE: 03-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/947,263
 FILING DATE: 18-SEP-1992

ATTORNEY/AGENT INFORMATION:
 NAME: Richard W. Bork
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4032US4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:
 LENGTH: 33 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-809-523-62

Query Match 64.7%; Score 19.4; DB 3; Length 33;
 Best Local Similarity 79.3%; Pred. No. 46;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
 |||||
 Db 32 CTGGTCTTTATGCGCCCTCGAATTC 4

RESULT 7

US-08-471-971-62/c
 ; Sequence 62, Application US/08471971
 ; Patent No. 6287759

GENERAL INFORMATION:
 APPLICANT: Tsarev, Sergei. A., Emerson,
 APPLICANT: Suzanne U., Purcell, Robert H.
 TITLE OF INVENTION: Recombinant Proteins Of
 TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
 TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 NUMBER OF SEQUENCES: 107
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,971
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US08/316,765
 FILING DATE: 03-OCT-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US07/947,263
 FILING DATE: 18-SEP-1992

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Richard W. Bork
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4032US2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 33 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-471-971-62

Query Match 64.7%; Score 19.4; DB 3; Length 33;
 Best Local Similarity 79.3%; Pred. No. 46;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
 |||||
 Db 32 CTGGTCTTTATGCGCCCTCGAATTC 4

RESULT 8

US-09-402-776-62/c
 ; Sequence 62, Application US/09402776
 ; Patent No. 6458562

GENERAL INFORMATION:
 APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
 APPLICANT: Tsarev, Sergei. A., and Robinson, Robin A.
 TITLE OF INVENTION: Recombinant Proteins Of
 TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
 TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,776
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/840,316
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4255
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-402-776-62

Query Match 64.7%; Score 19.4; DB 3; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGCCCGCCCTTGAATTCC 29
|||||
DB 32 CTGGTTCTTTATGCGCCCGCTCGAATTCC 4

RESULT 9
US-08-470-246-62/c
Sequence 62, Application US/08470246
Patent No. 6696242
GENERAL INFORMATION:
APPLICANT: Tearev, Sergei. A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,246
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US08/316,765
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US07/947,263
FILING DATE: 18-SEP-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4032US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-470-246-62

Query Match 64.7%; Score 19.4; DB 4; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCTTGAATTCC 29
|||||
DB 32 CTGGTTCTTTATGCGCCCGCTCGAATTCC 4

RESULT 10
US-08-316-765-62/c
Sequence 62, Application US/08316765
Patent No. 6706873
GENERAL INFORMATION:
APPLICANT: Tearev, Sergei. A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,765
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US07/947,263
FILING DATE: 18-SEP-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4032US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-316-765-62

```
Query Match          64.7%; Score 19.4; DB 4; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CTGGATCCTTGCCTCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 32 CTGGTTCTTTATGCGCGCCCTCGAATTC 4

RESULT 11
US-09-724-475-62/c
; Sequence 62, Application US/09724475
; Patent No. 6787145
; GENERAL INFORMATION:
; APPLICANT: Tearev, Sergei. A., Emerson,
; Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; A Pakistani Strain Of Hepatitis E And Their
; Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,475
; FILING DATE: 28-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,523
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US08/316,765
; FILING DATE: 03-OCT-1994
; APPLICATION NUMBER: 07/947,263
; FILING DATE: 18-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard W. Bork
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032US4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-09-724-475-62

Query Match          64.7%; Score 19.4; DB 4; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CTGGATCCTTGCCTCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 32 CTGGTTCTTTATGCGCGCCCTCGAATTC 4

RESULT 12
PCT-US93-08849A-62/c
; Sequence 62, Application PC/TUS9308849A
; GENERAL INFORMATION:
; APPLICANT: Tearev, Sergei A., Emerson,
; Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; A Pakistani Strain Of Hepatitis E And Their
; Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08849
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PCT-US93-08849A-62/c

Query Match          64.7%; Score 19.4; DB 5; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CTGGATCCTTGCCTCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 32 CTGGTTCTTTATGCGCGCCCTCGAATTC 4

RESULT 13
PCT-US93-08849-62/c
; Sequence 62, Application PC/TUS9308849
; GENERAL INFORMATION:
; APPLICANT: Tearev, Sergei A., Emerson,
; Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; A Pakistani Strain Of Hepatitis E And Their
; Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08849
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PCT-US93-08849A-62/c

Query Match          64.7%; Score 19.4; DB 5; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CTGGATCCTTGCCTCCCGCCCTTGAATTC 29
    ||||| ||||| ||||| ||||| |||||
Db 32 CTGGTTCTTTATGCGCGCCCTCGAATTC 4
```

; APPLICATION NUMBER: 07/947,263
; FILING DATE: 18-SEP-1992
; NAME:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bork, Richard, W.
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US93-08849-62

Query Match 64.7%; Score 19.4; DB 5; Length 33;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 32 CTGGTCTTTATGCCCGCCCTCGAATTC 4

RESULT 14
US-09-949-016-14957/c
; Sequence 14957, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14957
; LENGTH: 325034
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14957

Query Match 64.7%; Score 19.4; DB 4; Length 325034;
Best Local Similarity 79.3%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 308900 CTGGATCCTTGGCCCGGAACCTGCATTC 308872

RESULT 15
US-09-949-016-11774/c
; Sequence 11774, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11774
; LENGTH: 389504
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11774

Query Match 64.7%; Score 19.4; DB 4; Length 389504;
Best Local Similarity 79.3%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 325370 CTGGATCCTTGGCCCGGAACCTGCATTC 325342

RESULT 16
US-09-949-016-12832
; Sequence 12832, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12832
; LENGTH: 37711
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12832

Query Match 62.7%; Score 18.8; DB 4; Length 37711;
Best Local Similarity 76.7%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 30
DB 9585 CTGATCTTGGCCACCATCTGTAATTC 9614

RESULT 17
US-09-949-016-16704
; Sequence 16704, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16704
; LENGTH: 37712
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16704

Query Match          62.7%; Score 18.8; DB 4; Length 37712;
Best Local Similarity 76.7%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCGCCCTTGAATTC 30
   |||||||
Db 9585 CTGATCCTTGCCACACCATCTGTAATTC 9614

RESULT 18
US-08-247-901C-1/c
; Sequence 1, Application US/08247901C
; Patent No. 5750384
; GENERAL INFORMATION:
; APPLICANT: Jacobs et al
; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,901C
; FILING DATE: May 23, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/057,531
; FILING DATE: April 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Elizabeth A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/273
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50341
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: L5 shuttle phasmid sequence.
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: L5 mycobacteriophage
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:

; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: No. 5750384e
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1

Query Match          62.7%; Score 18.8; DB 1; Length 50341;
Best Local Similarity 90.9%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGATCCTTGCCCGCCCGCCCTTGA 24
   |||||||
Db 7081 GGATCCTTGCCCGCCCGCCCTTGA 7060

RESULT 19
US-09-075-904-1/c
; Sequence 1, Application US/09075904
; Patent No. 5994137
; GENERAL INFORMATION:
; APPLICANT: Jacobs, et al.
; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,904
; FILING DATE: May 11, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/247,901
; FILING DATE: May 23, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Elizabeth A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/475
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50341
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: L5 shuttle phasmid sequence
```



```

; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: L5 mycobacteriophage
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: No. 5994137e
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
;
US-09-075-904-1
Query Match 62.7%; Score 18.8; DB 2; Length 50341;
Best Local Similarity 90.9%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGATCCTTGGCCCGCCCTTGA 24
Db 7081 GGATCCTTGGCCCGCCCTTGA 7060

RESULT 20
US-09-426-436-1/c
; Sequence 1, Application US/09426436
; Patent No. 6225066
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/426,436
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,557
; FILING DATE:

```

```

; APPLICATION NUMBER: US/08/057,531
; FILING DATE:
; APPLICATION NUMBER: 07/833,431
; FILING DATE: February 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: 96700/238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52297
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: phage genome sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: not applicable.
; ORIGINAL SOURCE:
; ORGANISM: mycobacteriophage L5
; STRAIN: not applicable
; INDIVIDUAL ISOLATE: L5
; DEVELOPMENTAL STAGE: not applicable
; HAPLOTYPE: not applicable
; TISSUE TYPE: not applicable
; CELL TYPE: not applicable
; CELL LINE: not applicable
; ORGANELLE: not applicable
; IMMEDIATE SOURCE: mycobacteriophage L5 particles
; POSITION IN GENOME: entire genome
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Hatfull and Sarkis
; TITLE: DNA Sequence, Structure and Gene
; TITLE: Expression of Mycobacteriophage L5:
; TITLE: A Phage System for Mycobacterial
; TITLE: Genetics
; JOURNAL: Molecular Microbiology
; VOLUME: 7
; PAGES: 395-405
; DATE: 1993
;
US-09-426-436-1
Query Match 62.7%; Score 18.8; DB 3; Length 52297;
Best Local Similarity 90.9%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGATCCTTGGCCCGCCCTTGA 24
Db 6978 GGATCCTTGGCCCGCCCTTGA 6957

RESULT 21
US-08-705-557-1/c
; Sequence 1, Application US/08705557
; Patent No. 6300061
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:

```

RESIST 24

3 GGA TCC TTG CCCC CCCC CTTGA 24

<p>US-09-949-016-16030</p> <p>; Sequence 16030, Application US/09949016</p> <p>; Patent No. 6812339</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: VENTER, J. Craig et al.</p> <p>; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>; FILE REFERENCE: CL001307</p> <p>; CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>; PRIOR FILING DATE: 2000-04-14</p> <p>; PRIOR APPLICATION NUMBER: 60/241,755</p> <p>; PRIOR FILING DATE: 2000-10-20</p> <p>; PRIOR APPLICATION NUMBER: 60/237,768</p> <p>; PRIOR FILING DATE: 2000-10-03</p> <p>; PRIOR APPLICATION NUMBER: 60/231,498</p> <p>; NUMBER OF SEQ ID NOS: 207012</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 16030</p> <p>; LENGTH: 12163</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Human</p> <p>US-09-949-016-16030</p>	<p>Query Match 61.3%; Score 18.4; DB 4; Length 12163;</p> <p>Best Local Similarity 95.0%; Pred. No. 2.5e+02;</p> <p>Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;</p>	<p>Qy 7 CCTTGCCCGCCCTTGAAT 26</p> <p>Db 1499 CCGTGCCCGCCCTTGAAT 1518</p>	<p>RESULT 25</p> <p>US-09-949-016-12202</p> <p>; Sequence 12202, Application US/09949016</p> <p>; Patent No. 6812339</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: VENTER, J. Craig et al.</p> <p>; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>; FILE REFERENCE: CL001307</p> <p>; CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>; PRIOR FILING DATE: 2000-04-14</p> <p>; PRIOR APPLICATION NUMBER: 60/241,755</p> <p>; PRIOR FILING DATE: 2000-10-20</p> <p>; PRIOR APPLICATION NUMBER: 60/237,768</p> <p>; PRIOR FILING DATE: 2000-10-03</p> <p>; PRIOR APPLICATION NUMBER: 60/231,498</p> <p>; NUMBER OF SEQ ID NOS: 207012</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 12202</p> <p>; LENGTH: 12164</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Human</p> <p>US-09-949-016-12202</p>	<p>Query Match 61.3%; Score 18.4; DB 4; Length 12164;</p> <p>Best Local Similarity 95.0%; Pred. No. 2.5e+02;</p> <p>Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;</p>	<p>Qy 7 CCTTGCCCGCCCTTGAAT 26</p> <p>Db 1499 CCGTGCCCGCCCTTGAAT 1518</p>	<p>RESULT 26</p> <p>US-09-949-016-13683/c</p> <p>; Sequence 13683, Application US/09949016</p> <p>; Patent No. 6812339</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: VENTER, J. Craig et al.</p> <p>; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>; FILE REFERENCE: CL001307</p> <p>; CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>; PRIOR FILING DATE: 2000-04-14</p> <p>; PRIOR APPLICATION NUMBER: 60/241,755</p> <p>; PRIOR FILING DATE: 2000-10-20</p> <p>; PRIOR APPLICATION NUMBER: 60/237,768</p> <p>; PRIOR FILING DATE: 2000-10-03</p> <p>; PRIOR APPLICATION NUMBER: 60/231,498</p> <p>; NUMBER OF SEQ ID NOS: 207012</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 13683</p> <p>; LENGTH: 13426</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Human</p> <p>US-09-949-016-13683</p>	<p>Query Match 61.3%; Score 18.4; DB 4; Length 13426;</p> <p>Best Local Similarity 78.6%; Pred. No. 2.5e+02;</p> <p>Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;</p>	<p>Qy 3 GGATCCTTCGCCCGCCCTTGAATTCCC 30</p> <p>Db 12138 GGATCCTTCACCGCCCGCCAGCATATCCC 12111</p>	<p>RESULT 27</p> <p>US-09-949-016-15332</p> <p>; Sequence 15332, Application US/09949016</p> <p>; Patent No. 6812339</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: VENTER, J. Craig et al.</p> <p>; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>; FILE REFERENCE: CL001307</p> <p>; CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>; PRIOR FILING DATE: 2000-04-14</p> <p>; PRIOR APPLICATION NUMBER: 60/241,755</p> <p>; PRIOR FILING DATE: 2000-10-20</p> <p>; PRIOR APPLICATION NUMBER: 60/237,768</p> <p>; PRIOR FILING DATE: 2000-10-03</p> <p>; PRIOR APPLICATION NUMBER: 60/231,498</p> <p>; NUMBER OF SEQ ID NOS: 207012</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 15332</p> <p>; LENGTH: 71251</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Human</p> <p>US-09-949-016-15332</p>	<p>Query Match 61.3%; Score 18.4; DB 4; Length 71251;</p> <p>Best Local Similarity 78.6%; Pred. No. 3.1e+02;</p> <p>Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;</p>	<p>Qy 3 GGATCCTTCGCCCGCCCTTGAATTCCC 30</p> <p>Db 60499 GCCTTCCTGCCCCCTCCCTTCATTTACC 60526</p>	<p>RESULT 28</p> <p>US-09-949-016-12733/c</p> <p>; Sequence 12733, Application US/09949016</p> <p>; Patent No. 6812339</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: VENTER, J. Craig et al.</p> <p>; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>; FILE REFERENCE: CL001307</p> <p>; CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>; PRIOR FILING DATE: 2000-04-14</p> <p>; PRIOR APPLICATION NUMBER: 60/241,755</p> <p>; PRIOR FILING DATE: 2000-10-20</p> <p>; PRIOR APPLICATION NUMBER: 60/237,768</p> <p>; PRIOR FILING DATE: 2000-10-03</p> <p>; PRIOR APPLICATION NUMBER: 60/231,498</p> <p>; NUMBER OF SEQ ID NOS: 207012</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 12733</p> <p>; LENGTH: 71251</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Human</p> <p>US-09-949-016-12733</p>	<p>Query Match 61.3%; Score 18.4; DB 4; Length 71251;</p> <p>Best Local Similarity 78.6%; Pred. No. 3.1e+02;</p> <p>Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;</p>	<p>Qy 3 GGATCCTTCGCCCGCCCTTGAATTCCC 30</p> <p>Db 60499 GCCTTCCTGCCCCCTCCCTTCATTTACC 60526</p>
--	--	--	---	--	--	---	--	--	---	--	---	---	--	---

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12733
; LENGTH: 178883
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12733

Query Match          61.3%; Score 18.4; DB 4; Length 178883;
Best Local Similarity 78.6%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 28
DB 123968 CTGGATGCTGGCCCGCCACCTGAATGC 123941

RESULT 29
US-09-949-016-13039/c
; Sequence 13039, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13039
; LENGTH: 178884
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13039

Query Match          61.3%; Score 18.4; DB 4; Length 178884;
Best Local Similarity 78.6%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 28
DB 123968 CTGGATGCTGGCCCGCCACCTGAATGC 123941

RESULT 30
US-09-949-016-15078/c
; Sequence 15078, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15078
; LENGTH: 114426
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15078

Query Match          60.7%; Score 18.2; DB 4; Length 114426;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 TCCTTGGCCCGCCCTTGAATTC 28
DB 28656 TCATTCCCGCCACCCCTTGAATTC 28634

RESULT 31
US-09-949-016-142664
; Sequence 142664, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 142664
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-142664

Query Match          60.0%; Score 18; DB 4; Length 601;
Best Local Similarity 80.8%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TGGATCTTGGCCCGCCCTTGAATT 27
DB 431 TGGGTCATTGCTCAGCCCTTGAGTT 456

RESULT 32
US-08-890-542A-1/c
; Sequence 1, Application US/08890542A
; Patent No. 5840509
; GENERAL INFORMATION:
; APPLICANT: Ni, Binhui
; APPLICANT: Paul, Steven M
; APPLICANT: Wu, Xin
; TITLE OF INVENTION: PROTEASE AND RELATED NUCLEIC ACID
; COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
```

```
;
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/890,542A
; FILING DATE: 09-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..906
;
US-08-890-542A-1

Query Match 60.0%; Score 18; DB 2; Length 1177;
Best Local Similarity 80.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 GATCCTTGGCCCGCCCTTGAATTC 29
DB 720 GAACCATGACCGTCCCTTGAATTC 695

RESULT 33
US-08-890-542A-3/c
; Sequence 3, Application US/08890542A
; Patent No. 5840509
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M
; APPLICANT: Wu, Xin
; TITLE OF INVENTION: PROTEASE AND RELATED NUCLEIC ACID
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/890,542A
; FILING DATE: 09-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
;
US-08-890-542A-3

Query Match 60.0%; Score 18; DB 2; Length 1177;
Best Local Similarity 80.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 GATCCTTGGCCCGCCCTTGAATTC 29
DB 720 GAACCATGACCGTCCCTTGAATTC 695

RESULT 34
US-09-949-016-15774
; Sequence 15774, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15774
; LENGTH: 43307
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-15774

Query Match 60.0%; Score 18; DB 4; Length 43307;
Best Local Similarity 80.8%; Pred. No. 4.2e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 TGGATCCTTGGCCCGCCCTTGAATT 27
DB 32024 TGGTCATTGCTCAGCCCCCTTGAGTT 32049

RESULT 35
US-09-621-976-1833/c
; Sequence 1833, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1833
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..450
; NAME/KEY: sig_peptide
```


QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 118 TTGATCATTGGCACTCCCTATGAATGCC 146

RESULT 40

US-09-949-016-203251
; Sequence 203251, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203251
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203251

Query Match 59.3%; Score 17.8; DB 4; Length 601;
Best Local Similarity 75.9%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 390 TTGATCATTGGCACTCCCTATGAATGCC 418

RESULT 41

US-09-949-016-203252
; Sequence 203252, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203252
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203252

Query Match 59.3%; Score 17.8; DB 4; Length 601;
Best Local Similarity 75.9%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TGGATCCTTGGCCCGCCCTTGAATTC 30
DB 396 TTGATCATTGGCACTCCCTATGAATGCC 424

RESULT 42

US-09-949-016-12125
; Sequence 12125, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12125
; LENGTH: 20565
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12125

Query Match 59.3%; Score 17.8; DB 4; Length 20565;
Best Local Similarity 75.9%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 3286 CTCATCTTGTCCCGCCCTGGATGTC 3314

RESULT 43

US-09-949-016-13990
; Sequence 13990, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13990
; LENGTH: 20566
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13990

Query Match 59.3%; Score 17.8; DB 4; Length 20566;
Best Local Similarity 75.9%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
DB 3286 CTCATCTTGTCCCGCCCTGGATGTC 3314

RESULT 44

US-09-949-016-12392
; Sequence 12392, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12392
; LENGTH: 23928
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12392

Query Match 59.3%; Score 17.8; DB 4; Length 23928;
Best Local Similarity 75.9%; Pred. No. 4.8e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCTTGAATTCC 29
DB 475 CCGGCTCCATCCCGCGCCCTAGAAATTCG 503

RESULT 45

US-09-949-016-16315
; Sequence 16315, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16315
; LENGTH: 23928
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16315

Query Match 59.3%; Score 17.8; DB 4; Length 23928;
Best Local Similarity 75.9%; Pred. No. 4.8e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCTTGAATTCC 29
DB 475 CCGGCTCCATCCCGCGCCCTAGAAATTCG 503

RESULT 46

US-09-949-016-15959/c
; Sequence 15959, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15959
; LENGTH: 26115
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15959

Query Match 59.3%; Score 17.8; DB 4; Length 26115;
Best Local Similarity 75.9%; Pred. No. 4.9e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TGGATCCTTGCCCGCCCTTGAATTCCC 30
DB 9617 TGGTCTTGGCACCAGCCCGGACTTGCC 9589

RESULT 47

US-09-949-016-17477/c
; Sequence 17477, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17477
; LENGTH: 27589
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17477

Query Match 59.3%; Score 17.8; DB 4; Length 27589;
Best Local Similarity 75.9%; Pred. No. 4.9e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TGGATCCTTGCCCGCCCTTGAATTCCC 30
DB 20519 TTGATCATGGCACTCCCTATGAATTGCC 20491

RESULT 48

US-09-949-016-15973
; Sequence 15973, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03


```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15973
; LENGTH: 37254
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(37254)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15973

Query Match          59.3%; Score 17.8; DB 4; Length 37254;
Best Local Similarity 75.9%; Pred. No. 5.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      2 TGGATCCTTGGCCCGCCCTTGAATTCC 30
      ||| ||||| ||||| ||||| |||||
Db      27948 TGGTCCCTTGGCCCTGGGCTTGGTTTCCC 27976

RESULT 49
US-09-949-016-16477
; Sequence 16477, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16477
; LENGTH: 72549
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(72549)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16477

Query Match          59.3%; Score 17.8; DB 4; Length 72549;
Best Local Similarity 90.5%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CTGGATCCTTGGCCCGCCCT 21
      ||| ||||| ||||| |||||
Db      59740 CTGGTCTTGCACCGCCCT 59760

RESULT 50
US-09-949-016-17078/c
; Sequence 17078, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17078
; LENGTH: 123463
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(123463)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17078

Query Match          59.3%; Score 17.8; DB 4; Length 123463;
Best Local Similarity 75.9%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 CTGGATCCTTGGCCCGCCCTTGAATTCC 29
      ||| ||||| ||||| ||||| |||||
Db      71574 CTGAATCCTGGCTCCACCACTTGACTGCC 71546

Search completed: April 25, 2005, 15:09:44
Job time : 124 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:16:14 ; Search time 374.483 Seconds
(without alignments)
486.355 Million cell updates/sec

Title: US-10-010-476-10

Perfect score: 30

Sequence: 1 CTGGATCTTGGCCCGCCCTTGATTCCTC 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	30	100.0	30	15	US-10-010-476-9
C 2	30	100.0	30	15	US-10-010-476-10
C 3	30	100.0	50	15	US-10-010-476-13
C 4	22	73.3	22	15	US-10-010-476-14
C 5	21.6	72.0	4562	18	US-10-664-705-74
C 6	21	70.0	593	18	US-10-425-115-93962
C 7	20.4	68.0	424	14	US-10-116-802-513
C 8	20	66.7	574	18	US-10-425-115-170110
C 9	20	66.7	1203	18	US-10-650-112-27
C 10	20	66.7	1760	17	US-10-152-319A-1814
C 11	19.8	66.0	456	9	US-09-783-590-3630
C 12	19.6	65.3	4385	17	US-10-291-265-476
C 13	19.6	65.3	50000	17	US-10-364-505-9
C 14	19.6	65.3	50000	18	US-10-681-199-9
C 15	19.6	65.3	191597	18	US-10-719-993-6802
C 16	19.4	64.7	540	18	US-10-430-201-2349
C 17	19.4	64.7	540	18	US-10-430-201-2350
C 18	19.4	64.7	1041	17	US-10-369-433-25548
C 19	19.4	64.7	96593	11	US-09-997-722-151
C 20	19.2	64.0	1189	18	US-10-723-860-7320
C 21	19	63.3	371	18	US-10-425-115-97766
C 22	19	63.3	741	18	US-10-437-963-24928
C 23	18.8	62.7	30	15	US-10-010-476-11
C 24	18.8	62.7	30	15	US-10-010-476-12
C 25	18.8	62.7	577	16	US-10-029-386-13283
C 26	18.8	62.7	680	13	US-10-027-632-28086
C 27	18.8	62.7	680	17	US-10-027-632-28086
C 28	18.8	62.7	708	14	US-10-198-846-596
C 29	18.8	62.7	816	13	US-10-027-632-167839
C 30	18.8	62.7	816	17	US-10-027-632-167839
C 31	18.8	62.7	1051	11	US-09-876-143-1396
C 32	18.8	62.7	2021	17	US-10-104-047-449
C 33	18.8	62.7	2048	17	US-10-094-749-137
C 34	18.8	62.7	2524	18	US-10-335-053-162
C 35	18.8	62.7	2524	18	US-10-437-963-77151
C 36	18.8	62.7	9370	10	US-09-764-891-9763
C 37	18.8	62.7	21328	18	US-10-719-993-6908
C 38	18.8	62.7	51198	13	US-10-087-192-1729
C 39	18.8	62.7	786452	18	US-10-719-993-6822
C 40	18.6	62.0	180	17	US-10-242-535A-6953
C 41	18.6	62.0	180	17	US-10-085-783A-6953
C 42	18.6	62.0	262	17	US-10-242-535A-42504
C 43	18.6	62.0	262	17	US-10-085-783A-42504
C 44	18.6	62.0	277	17	US-10-242-535A-9468
C 45	18.6	62.0	277	17	US-10-085-783A-9468
C 46	18.6	62.0	505	18	US-10-425-115-86331
C 47	18.6	62.0	646	13	US-10-027-632-7977
C 48	18.6	62.0	646	13	US-10-027-632-7978
C 49	18.6	62.0	646	17	US-10-027-632-7977
C 50	18.6	62.0	646	17	US-10-027-632-7978
C 51	18.6	62.0	742	13	US-10-027-632-98673
C 52	18.6	62.0	742	17	US-10-027-632-98673
C 53	18.6	62.0	1789	17	US-10-172-118-585
C 54	18.6	62.0	1789	17	US-10-342-887-585
C 55	18.6	62.0	21925	15	US-10-017-161-2415
C 56	18.6	62.0	21925	15	US-10-292-798-2055
C 57	18.6	62.0	26184	15	US-10-017-161-2043
C 58	18.6	62.0	26184	17	US-10-292-798-1689
C 59	18.6	62.0	37644	14	US-10-043-160-33
C 60	18.6	62.0	96599	14	US-10-105-948-4
C 61	18.6	62.0	96599	17	US-10-052-482-178
C 62	18.6	62.0	158001	17	US-10-211-179-11
C 63	18.6	62.0	378361	10	US-09-901-136-3
C 64	18.6	62.0	378361	18	US-10-483-329-3
C 65	18.4	61.3	271	17	US-10-424-599-77005
C 66	18.4	61.3	332	18	US-10-425-115-155202
C 67	18.4	61.3	421	13	US-10-027-632-49848
C 68	18.4	61.3	421	17	US-10-027-632-49848
C 69	18.4	61.3	509	17	US-10-424-599-125632
C 70	18.4	61.3	628	13	US-10-027-632-246670
C 71	18.4	61.3	628	17	US-10-027-632-246670
C 72	18.4	61.3	647	17	US-10-264-049-1101
C 73	18.4	61.3	2066	17	US-10-074-024-533
C 74	18.4	61.3	2391	15	US-10-128-714-1403
C 75	18.4	61.3	2391	15	US-10-128-714-2403
C 76	18.4	61.3	2391	15	US-10-128-714-5403
C 77	18.4	61.3	2391	15	US-10-128-714-7403
C 78	18.4	61.3	2925	18	US-10-437-963-79900
C 79	18.4	61.3	4390	15	US-10-128-714-403
C 80	18.4	61.3	4390	15	US-10-128-714-5403
C 81	18.4	61.3	5997	17	US-10-074-024-535
C 82	18.4	61.3	6093	17	US-10-074-024-535
C 83	18.4	61.3	94001	17	US-10-210-838-20
C 84	18.4	61.3	194883	13	US-10-087-192-826

Sequence 476, App
Sequence 9, Appli
Sequence 9, Appli
Sequence 6802, Ap
Sequence 2349, Ap
Sequence 2350, Ap
Sequence 25548, A
Sequence 151, App
Sequence 97766, A
Sequence 24928, A
Sequence 11, Appl
Sequence 12, Appl
Sequence 13883, A
Sequence 28086, A
Sequence 28086, A
Sequence 596, App
Sequence 167839,
Sequence 167839,
Sequence 1396, Ap
Sequence 449, App
Sequence 137, App
Sequence 162, App
Sequence 77151, A
Sequence 9763, Ap
Sequence 6908, Ap
Sequence 1729, Ap
Sequence 6822, Ap
Sequence 6953, Ap
Sequence 6953, Ap
Sequence 42504, A
Sequence 42504, A
Sequence 9468, Ap
Sequence 9468, Ap
Sequence 86331, A
Sequence 7977, Ap
Sequence 7978, Ap
Sequence 7977, Ap
Sequence 7978, Ap
Sequence 98673, A
Sequence 98673, A
Sequence 585, App
Sequence 585, App
Sequence 2415, Ap
Sequence 2055, Ap
Sequence 2043, Ap
Sequence 1689, Ap
Sequence 33, Appl
Sequence 4, Appl
GENERAL INFORMATI
Sequence 3, Appli
Sequence 3, Appli
Sequence 77005, A
Sequence 155202,
Sequence 49848, A
Sequence 49848, A
Sequence 125632,
Sequence 246670,
Sequence 246670,
Sequence 1101, Ap
Sequence 533, App
Sequence 1403, Ap
Sequence 2403, Ap
Sequence 6403, Ap
Sequence 7403, Ap
Sequence 79900, A
Sequence 403, App
Sequence 5403, Ap
Sequence 534, App
Sequence 535, App
Sequence 20, Appl
Sequence 826, App

c 85	18.4	61.3	1691139	14	US-10-067-514-1	Sequence 1, Appli	c 158	17.8	59.3	111836	18	US-10-322-281-51	Sequence 51, Appl
c 86	18.4	61.3	1691139	17	US-10-419-723-1	Sequence 1, Appli	c 159	17.8	59.3	183651	19	US-10-484-577-664	Sequence 664, App
c 87	18.2	60.7	445	17	US-10-424-599-5021	Sequence 5021, Ap	c 160	17.8	59.3	186591	19	US-10-484-577-682	Sequence 682, App
c 88	18.2	60.7	577	13	US-10-027-632-133790	Sequence 133790,	c 161	17.8	59.3	203132	18	US-10-322-281-459	Sequence 459, App
c 89	18.2	60.7	577	17	US-10-027-632-133790	Sequence 133790,	c 162	17.8	59.3	208648	19	US-10-484-577-663	Sequence 663, App
c 90	18.2	60.7	598	13	US-10-027-632-91173	Sequence 91173, A	c 163	17.8	59.3	208648	19	US-10-484-577-668	Sequence 668, App
c 91	18.2	60.7	598	13	US-10-027-632-91173	Sequence 91174, A	c 164	17.8	59.3	561515	18	US-10-741-601-5682	Sequence 5682, Ap
c 92	18.2	60.7	598	13	US-10-027-632-304408	Sequence 304408,	c 165	17.8	59.3	561515	18	US-10-741-601-17730	Sequence 17730, A
c 93	18.2	60.7	598	13	US-10-027-632-304409	Sequence 304409,	c 166	17.6	58.7	201	18	US-10-719-993-22882	Sequence 22882, A
c 94	18.2	60.7	598	17	US-10-027-632-91173	Sequence 91173, A	c 167	17.6	58.7	201	18	US-10-719-993-42715	Sequence 42715, A
c 95	18.2	60.7	598	17	US-10-027-632-91174	Sequence 91174, A	c 168	17.6	58.7	223	17	US-10-424-599-68929	Sequence 68929, A
c 96	18.2	60.7	598	17	US-10-027-632-304408	Sequence 304408,	c 169	17.6	58.7	255	17	US-10-424-599-141190	Sequence 141190,
c 97	18.2	60.7	598	17	US-10-027-632-304409	Sequence 304409,	c 170	17.6	58.7	333	9	US-09-864-761-31553	Sequence 31553, A
c 98	18.2	60.7	637	13	US-10-027-632-133791	Sequence 133791,	c 171	17.6	58.7	366	18	US-10-425-115-35066	Sequence 35066, A
c 99	18.2	60.7	637	17	US-10-027-632-133791	Sequence 133791,	c 172	17.6	58.7	396	9	US-09-560-863-724	Sequence 724, App
c 100	18.2	60.7	716	17	US-10-258-666-29	Sequence 29, Appl	c 173	17.6	58.7	468	18	US-10-425-115-25844	Sequence 25844, A
c 101	18.2	60.7	787	17	US-10-258-666-30	Sequence 30, Appl	c 174	17.6	58.7	486	10	US-09-918-995-31331	Sequence 31331, A
c 102	18.2	60.7	2298	17	US-10-108-260A-1375	Sequence 1375, Ap	c 175	17.6	58.7	498	9	US-09-864-761-15021	Sequence 15021, A
c 103	18.2	60.7	2568	17	US-10-312-088-12	Sequence 12, Appl	c 176	17.6	58.7	561	18	US-10-437-963-72709	Sequence 72709, A
c 104	18.2	60.7	2622	15	US-10-176-306-81	Sequence 81, Appl	c 177	17.6	58.7	578	16	US-10-029-386-13249	Sequence 13249, A
c 105	18.2	60.7	2634	15	US-10-176-306-79	Sequence 79, Appl	c 178	17.6	58.7	676	17	US-10-424-599-87907	Sequence 87907, A
c 106	18.2	60.7	2867	17	US-10-312-088-11	Sequence 11, Appl	c 179	17.6	58.7	757	18	US-10-425-115-164041	Sequence 164041,
c 107	18.2	60.7	3657	14	US-10-198-846-14014	Sequence 14014, A	c 180	17.6	58.7	837	17	US-10-369-493-42868	Sequence 42868, A
c 108	18.2	60.7	5998	18	US-10-476-397-16	Sequence 16, Appl	c 181	17.6	58.7	1066	17	US-10-424-599-64356	Sequence 64356, A
c 109	18.2	60.7	6591	18	US-10-723-860-7162	Sequence 7162, Ap	c 182	17.6	58.7	1482	17	US-10-425-114-29519	Sequence 29519, A
c 110	18.2	60.7	33916	18	US-10-322-281-861	Sequence 861, App	c 183	17.6	58.7	1729	17	US-10-369-493-43180	Sequence 43180, A
c 111	18.2	60.7	251199	19	US-10-482-029-205	Sequence 205, App	c 184	17.6	58.7	1742	17	US-10-424-599-56276	Sequence 56276, A
c 112	18	60.0	80	17	US-10-384-245-1010	Sequence 1010, Ap	c 185	17.6	58.7	2347	14	US-10-163-547-2	Sequence 2, Appli
c 113	18	60.0	175	17	US-10-242-535A-4993	Sequence 4993, Ap	c 186	17.6	58.7	3174	18	US-10-425-115-143423	Sequence 143423,
c 114	18	60.0	175	17	US-10-085-738A-4993	Sequence 4993, Ap	c 187	17.6	58.7	3951	17	US-10-379-981-8	Sequence 8, Appli
c 115	18	60.0	366	18	US-10-425-115-98690	Sequence 98690, A	c 188	17.6	58.7	4483	17	US-10-302-172-390	Sequence 390, App
c 116	18	60.0	516	18	US-10-437-963-64252	Sequence 64252, A	c 189	17.6	58.7	4659	17	US-10-399-645-22	Sequence 22, Appl
c 117	18	60.0	542	13	US-10-027-632-91110	Sequence 91110, A	c 190	17.6	58.7	6960	15	US-10-311-455-2371	Sequence 2371, Ap
c 118	18	60.0	542	13	US-10-027-632-91111	Sequence 91111, A	c 191	17.6	58.7	48656	18	US-10-719-993-6910	Sequence 6910, Ap
c 119	18	60.0	542	13	US-10-027-632-304378	Sequence 304378,	c 192	17.6	58.7	48656	19	US-10-741-600-17829	Sequence 17829, A
c 120	18	60.0	542	13	US-10-027-632-304379	Sequence 304379,	c 193	17.6	58.7	389939	13	US-10-417-375-4	Sequence 4, Appli
c 121	18	60.0	542	17	US-10-027-632-91110	Sequence 91110, A	c 194	17.6	58.7	389182	18	US-10-087-192-1102	Sequence 1102, Ap
c 122	18	60.0	542	17	US-10-027-632-91131	Sequence 91111, A	c 195	17.6	58.7	1980090	19	US-10-719-993-6815	Sequence 6815, Ap
c 123	18	60.0	542	17	US-10-027-632-304378	Sequence 304378,	c 196	17.6	58.7	1980090	18	US-10-741-600-17676	Sequence 17676, A
c 124	18	60.0	542	17	US-10-027-632-304379	Sequence 304379,	c 197	17.4	58.0	201	19	US-10-741-600-72568	Sequence 72568, A
c 125	18	60.0	571	13	US-10-027-632-135325	Sequence 135325,	c 198	17.4	58.0	307	18	US-10-425-115-54935	Sequence 54935, A
c 126	18	60.0	571	17	US-10-027-632-135325	Sequence 135325,	c 199	17.4	58.0	309	18	US-10-425-115-118681	Sequence 118681,
c 127	18	60.0	710	13	US-10-027-632-101456	Sequence 101456,	c 200	17.4	58.0	327	9	US-09-764-869-1321	Sequence 1321, Ap
c 128	18	60.0	710	13	US-10-027-632-101456	Sequence 101456,	c 201	17.4	58.0	327	9	US-09-764-869-2242	Sequence 2242, Ap
c 129	18	60.0	1081	17	US-10-388-934-547	Sequence 547, App	c 202	17.4	58.0	327	14	US-10-091-504-1321	Sequence 1321, Ap
c 130	18	60.0	1081	19	US-10-870-387-13	Sequence 13, Appl	c 203	17.4	58.0	327	14	US-10-091-504-2242	Sequence 2242, Ap
c 131	18	60.0	2573	19	US-10-152-319A-2018	Sequence 2018, Ap	c 204	17.4	58.0	327	17	US-10-227-577-1321	Sequence 1321, Ap
c 132	18	60.0	80077	13	US-10-087-132-1789	Sequence 1789, Ap	c 205	17.4	58.0	327	17	US-10-227-577-2242	Sequence 2242, Ap
c 133	18	60.0	117289	18	US-10-719-993-6859	Sequence 6859, Ap	c 206	17.4	58.0	334	17	US-10-424-599-131207	Sequence 131207,
c 134	18	60.0	2256646	18	US-10-470-565-1	Sequence 1, Appli	c 207	17.4	58.0	345	18	US-10-437-963-71537	Sequence 71537, A
c 135	17.8	59.3	294	17	US-10-424-599-17197	Sequence 31797, A	c 208	17.4	58.0	377	18	US-10-425-115-3442	Sequence 3442, Ap
c 136	17.8	59.3	411	17	US-10-424-599-123414	Sequence 123414, A	c 209	17.4	58.0	450	17	US-10-424-599-103639	Sequence 103639,
c 137	17.8	59.3	458	18	US-10-425-115-87766	Sequence 87766, A	c 210	17.4	58.0	500	17	US-09-918-995-28941	Sequence 28941, A
c 138	17.8	59.3	461	10	US-09-814-353-5737	Sequence 5737, Ap	c 211	17.4	58.0	511	17	US-10-424-599-69152	Sequence 69152, A
c 139	17.8	59.3	461	10	US-09-814-353-12020	Sequence 12020, A	c 212	17.4	58.0	526	18	US-10-425-115-131927	Sequence 131927,
c 140	17.8	59.3	513	17	US-10-424-599-28550	Sequence 28550, A	c 213	17.4	58.0	571	10	US-09-918-995-28231	Sequence 28231, A
c 141	17.8	59.3	643	13	US-10-027-632-285387	Sequence 285387,	c 214	17.4	58.0	604	18	US-10-425-115-13100	Sequence 13100, A
c 142	17.8	59.3	643	17	US-10-027-632-285387	Sequence 285387,	c 215	17.4	58.0	717	17	US-10-320-797-2047	Sequence 2047, Ap
c 143	17.8	59.3	787	13	US-10-027-632-165831	Sequence 165831,	c 216	17.4	58.0	722	17	US-10-425-114-35707	Sequence 35707, A
c 144	17.8	59.3	787	17	US-10-027-632-165831	Sequence 165831,	c 217	17.4	58.0	778	18	US-10-425-115-46137	Sequence 46137, A
c 145	17.8	59.3	801	18	US-10-437-963-74918	Sequence 74918, A	c 218	17.4	58.0	784	18	US-10-363-345A-34635	Sequence 34635, A
c 146	17.8	59.3	867	18	US-10-425-115-145943	Sequence 145943, A	c 219	17.4	58.0	784	19	US-10-363-345A-34635	Sequence 34635, A
c 147	17.8	59.3	1254	17	US-10-425-114-15662	Sequence 15662, A	c 220	17.4	58.0	784	19	US-10-363-483A-34635	Sequence 34635, A
c 148	17.8	59.3	1305	18	US-10-425-115-834	Sequence 834, App	c 221	17.4	58.0	784	19	US-10-363-483A-34635	Sequence 34635, A
c 149	17.8	59.3	3026	13	US-10-027-632-115170	Sequence 115170,	c 222	17.4	58.0	800	13	US-10-027-632-155300	Sequence 155300,
c 150	17.8	59.3	3026	17	US-10-027-632-115170	Sequence 115170,	c 223	17.4	58.0	800	17	US-10-029-386-22976	Sequence 22976, A
c 151	17.8	59.3	3846	17	US-10-094-749-1088	Sequence 1088, Ap	c 224	17.4	58.0	823	16	US-10-029-386-22976	Sequence 22976, A
c 152	17.8	59.3	38146	18	US-10-719-993-6860	Sequence 6860, Ap	c 225	17.4	58.0	869	18	US-10-363-345A-35325	Sequence 35325, A
c 153	17.8	59.3	52128	18	US-10-719-993-7051	Sequence 7051, Ap	c 226	17.4	58.0	869	19	US-10-363-345A-35326	Sequence 35326, A
c 154	17.8	59.3	65423	18	US-10-367-094-32	Sequence 32, Appl	c 227	17.4	58.0	869	19	US-10-363-483A-35326	Sequence 35326, A
c 155	17.8	59.3	73995	13	US-10-087-192-208	Sequence 208, App	c 228	17.4	58.0	869	19	US-10-363-483A-35326	Sequence 35326, A
c 156	17.8	59.3	98606	13	US-10-087-192-31	Sequence 31, Appl	c 229	17.4	58.0	935	18	US-10-767-701-5466	Sequence 5466, Ap
c 157	17.8	59.3	106315	13	US-10-087-192-178	Sequence 178, App	c 230	17.4	58.0	1002	17	US-10-320-797-1047	Sequence 1047, Ap

c 231	17.4	58.0	1024	13	US-10-202-193-29	Sequence 29, Appl	304	17.2	57.3	704	17	US-10-027-632-15994	Sequence 15994, A
c 232	17.4	58.0	1681	17	US-10-369-493-26954	Sequence 26954, A	c 305	17.2	57.3	714	13	US-10-027-632-17579	Sequence 17579, A
c 233	17.4	58.0	2091	9	US-09-817-647-22	Sequence 22, Appl	c 306	17.2	57.3	714	17	US-10-027-632-17579	Sequence 17579, A
c 234	17.4	58.0	2091	9	US-09-877-665-22	Sequence 22, Appl	c 307	17.2	57.3	762	17	US-10-369-493-24795	Sequence 24795, A
c 235	17.4	58.0	2091	13	US-10-136-573A-22	Sequence 22, Appl	c 308	17.2	57.3	783	13	US-10-027-632-172220	Sequence 172220, A
c 236	17.4	58.0	2091	14	US-10-215-862-22	Sequence 22, Appl	c 309	17.2	57.3	783	17	US-10-027-632-172220	Sequence 172220, A
c 237	17.4	58.0	2091	19	US-10-941-116-22	Sequence 22, Appl	c 310	17.2	57.3	799	18	US-10-437-963-30515	Sequence 30515, A
c 238	17.4	58.0	2502	9	US-09-817-647-5	Sequence 5, Appl	c 311	17.2	57.3	818	18	US-10-767-795-6234	Sequence 6234, Ap
c 239	17.4	58.0	2502	9	US-09-877-665-5	Sequence 5, Appl	c 312	17.2	57.3	832	17	US-10-425-114-14257	Sequence 14257, A
c 240	17.4	58.0	2502	13	US-10-136-573A-5	Sequence 5, Appl	c 313	17.2	57.3	1050	17	US-10-369-493-45643	Sequence 45643, A
c 241	17.4	58.0	2502	14	US-10-215-862-5	Sequence 5, Appl	c 314	17.2	57.3	1050	18	US-10-793-639-349	Sequence 349, App
c 242	17.4	58.0	2502	19	US-10-941-116-5	Sequence 5, Appl	c 315	17.2	57.3	1144	18	US-10-767-701-11598	Sequence 11598, A
c 243	17.4	58.0	3002	17	US-10-320-797-47	Sequence 47, Appl	c 316	17.2	57.3	1287	17	US-10-369-493-33611	Sequence 33611, A
c 244	17.4	58.0	3833	13	US-10-087-192-707	Sequence 707, App	c 317	17.2	57.3	1335	17	US-10-425-114-10905	Sequence 10905, A
c 245	17.4	58.0	3975	18	US-10-755-889-675	Sequence 675, App	c 318	17.2	57.3	1341	17	US-10-424-599-126934	Sequence 126934, A
c 246	17.4	58.0	4153	9	US-09-823-245A-455	Sequence 455, App	c 319	17.2	57.3	1513	18	US-10-425-115-111636	Sequence 111636, A
c 247	17.4	58.0	4834	17	US-10-369-493-27135	Sequence 27135, A	c 320	17.2	57.3	1567	17	US-10-282-122A-23246	Sequence 23246, A
c 248	17.4	58.0	4974	15	US-10-156-761-2487	Sequence 2487, Ap	c 321	17.2	57.3	1713	18	US-10-437-963-72391	Sequence 72391, A
c 249	17.4	58.0	5317	9	US-09-332-888-1	Sequence 1, Appl	c 322	17.2	57.3	1933	18	US-10-425-115-38896	Sequence 38896, A
c 250	17.4	58.0	20000	18	US-10-303-420-20	Sequence 20, Appl	c 323	17.2	57.3	2026	18	US-10-437-963-90691	Sequence 90691, A
c 251	17.4	58.0	23833	13	US-10-087-192-706	Sequence 706, App	c 324	17.2	57.3	2731	17	US-10-104-047-284	Sequence 284, App
c 252	17.4	58.0	31527	11	US-09-997-722-211	Sequence 211, App	c 325	17.2	57.3	2850	17	US-10-133-937-26	Sequence 26, Appl
c 253	17.4	58.0	45664	19	US-10-741-600-17881	Sequence 17881, A	c 326	17.2	57.3	2850	17	US-10-159-563-26	Sequence 26, Appl
c 254	17.4	58.0	125860	19	US-10-741-600-18002	Sequence 18002, A	c 327	17.2	57.3	3129	13	US-10-027-632-112234	Sequence 112234, A
c 255	17.4	58.0	256493	13	US-10-087-192-1000	Sequence 1000, Ap	c 328	17.2	57.3	3129	13	US-10-027-632-112235	Sequence 112235, A
c 256	17.4	58.0	493631	13	US-10-087-192-205	Sequence 205, App	c 329	17.2	57.3	3129	17	US-10-027-632-112234	Sequence 112234, A
c 257	17.4	58.0	3186778	13	US-10-027-632-174961	Sequence 174961, A	c 330	17.2	57.3	3129	17	US-10-027-632-112235	Sequence 112235, A
c 258	17.4	58.0	3186778	17	US-10-027-632-174961	Sequence 174961, A	c 331	17.2	57.3	3377	13	US-10-027-632-113850	Sequence 113850, A
c 259	17.4	58.0	9025608	15	US-10-156-761-1	Sequence 1, Appl	c 332	17.2	57.3	3377	13	US-10-027-632-113850	Sequence 113850, A
c 260	17.2	57.3	201	19	US-10-741-600-49843	Sequence 49843, A	c 333	17.2	57.3	3620	17	US-10-094-749-502	Sequence 902, App
c 261	17.2	57.3	324	11	US-09-864-408A-8817	Sequence 8817, Ap	c 334	17.2	57.3	4545	17	US-10-347-470A-29	Sequence 29, Appl
c 262	17.2	57.3	333	18	US-10-437-963-36115	Sequence 36115, A	c 335	17.2	57.3	4545	18	US-10-322-281-585	Sequence 585, App
c 263	17.2	57.3	353	18	US-10-425-115-107276	Sequence 107276, A	c 336	17.2	57.3	4563	13	US-08-946-374-259	Sequence 259, App
c 264	17.2	57.3	388	11	US-09-801-944B-94	Sequence 94, Appl	c 337	17.2	57.3	4563	13	US-10-052-586-311	Sequence 311, App
c 265	17.2	57.3	411	18	US-10-437-963-100068	Sequence 100068, A	c 338	17.2	57.3	4563	14	US-10-174-590-311	Sequence 311, App
c 266	17.2	57.3	422	9	US-09-864-761-16101	Sequence 16101, A	c 339	17.2	57.3	4563	14	US-10-176-758-311	Sequence 311, App
c 267	17.2	57.3	422	9	US-09-864-761-32594	Sequence 32594, A	c 340	17.2	57.3	4563	14	US-10-175-737-311	Sequence 311, App
c 268	17.2	57.3	439	18	US-10-425-115-156398	Sequence 156398, A	c 341	17.2	57.3	4563	14	US-10-174-581-311	Sequence 311, App
c 269	17.2	57.3	446	18	US-10-425-115-159210	Sequence 159210, A	c 342	17.2	57.3	4563	14	US-10-176-483-311	Sequence 311, App
c 270	17.2	57.3	452	18	US-10-425-115-150468	Sequence 150468, A	c 343	17.2	57.3	4563	14	US-10-176-749-311	Sequence 311, App
c 271	17.2	57.3	452	18	US-10-425-115-151221	Sequence 151221, A	c 344	17.2	57.3	4563	14	US-10-176-914-311	Sequence 311, App
c 272	17.2	57.3	455	10	US-09-918-995-34463	Sequence 34463, A	c 345	17.2	57.3	4563	14	US-10-176-915-311	Sequence 311, App
c 273	17.2	57.3	482	18	US-10-357-930-13565	Sequence 13565, A	c 346	17.2	57.3	4563	14	US-10-173-706-311	Sequence 311, App
c 274	17.2	57.3	532	18	US-10-425-115-108936	Sequence 108936, A	c 347	17.2	57.3	4563	14	US-10-175-738-311	Sequence 311, App
c 275	17.2	57.3	545	18	US-10-425-115-116739	Sequence 116739, A	c 348	17.2	57.3	4563	14	US-10-175-752-311	Sequence 311, App
c 276	17.2	57.3	563	18	US-10-425-115-75742	Sequence 75742, A	c 349	17.2	57.3	4563	14	US-10-176-482-311	Sequence 311, App
c 277	17.2	57.3	578	13	US-10-027-632-245620	Sequence 245620, A	c 350	17.2	57.3	4563	14	US-10-176-757-311	Sequence 311, App
c 278	17.2	57.3	578	13	US-10-027-632-245621	Sequence 245621, A	c 351	17.2	57.3	4563	14	US-10-176-913-311	Sequence 311, App
c 279	17.2	57.3	578	13	US-10-027-632-245622	Sequence 245622, A	c 352	17.2	57.3	4563	14	US-10-180-552-311	Sequence 311, App
c 280	17.2	57.3	578	17	US-10-027-632-245620	Sequence 245620, A	c 353	17.2	57.3	4563	14	US-10-180-557-311	Sequence 311, App
c 281	17.2	57.3	578	17	US-10-027-632-245621	Sequence 245621, A	c 354	17.2	57.3	4563	14	US-10-173-700-311	Sequence 311, App
c 282	17.2	57.3	578	17	US-10-027-632-245622	Sequence 245622, A	c 355	17.2	57.3	4563	14	US-10-174-572-311	Sequence 311, App
c 283	17.2	57.3	603	13	US-10-027-632-285608	Sequence 285608, A	c 356	17.2	57.3	4563	14	US-10-174-579-311	Sequence 311, App
c 284	17.2	57.3	603	17	US-10-027-632-285608	Sequence 285608, A	c 357	17.2	57.3	4563	14	US-10-174-582-311	Sequence 311, App
c 285	17.2	57.3	606	13	US-10-027-632-182868	Sequence 182868, A	c 358	17.2	57.3	4563	14	US-10-174-588-311	Sequence 311, App
c 286	17.2	57.3	606	13	US-10-027-632-182869	Sequence 182869, A	c 359	17.2	57.3	4563	14	US-10-175-739-311	Sequence 311, App
c 287	17.2	57.3	606	13	US-10-027-632-304434	Sequence 304434, A	c 360	17.2	57.3	4563	14	US-10-175-740-311	Sequence 311, App
c 288	17.2	57.3	606	17	US-10-027-632-182868	Sequence 182868, A	c 361	17.2	57.3	4563	14	US-10-175-743-311	Sequence 311, App
c 289	17.2	57.3	606	17	US-10-027-632-182869	Sequence 182869, A	c 362	17.2	57.3	4563	14	US-10-176-488-311	Sequence 311, App
c 290	17.2	57.3	606	17	US-10-027-632-304434	Sequence 304434, A	c 363	17.2	57.3	4563	14	US-10-176-492-311	Sequence 311, App
c 291	17.2	57.3	606	18	US-10-425-115-75744	Sequence 75744, A	c 364	17.2	57.3	4563	14	US-10-176-492-311	Sequence 311, App
c 292	17.2	57.3	607	13	US-10-027-632-111307	Sequence 111307, A	c 365	17.2	57.3	4563	14	US-10-176-750-311	Sequence 311, App
c 293	17.2	57.3	607	13	US-10-027-632-111308	Sequence 111308, A	c 366	17.2	57.3	4563	14	US-10-176-985-311	Sequence 311, App
c 294	17.2	57.3	607	17	US-10-027-632-111307	Sequence 111307, A	c 367	17.2	57.3	4563	14	US-10-176-987-311	Sequence 311, App
c 295	17.2	57.3	607	17	US-10-027-632-111308	Sequence 111308, A	c 368	17.2	57.3	4563	14	US-10-176-992-311	Sequence 311, App
c 296	17.2	57.3	608	13	US-10-027-632-93467	Sequence 93467, A	c 369	17.2	57.3	4563	14	US-10-176-993-311	Sequence 311, App
c 297	17.2	57.3	608	17	US-10-027-632-93467	Sequence 93467, A	c 370	17.2	57.3	4563	14	US-10-184-658-311	Sequence 311, App
c 298	17.2	57.3	608	18	US-10-425-115-14574	Sequence 14574, A	c 371	17.2	57.3	4563	14	US-10-176-991-311	Sequence 311, App
c 299	17.2	57.3	609	17	US-10-027-632-102153	Sequence 102153, A	c 372	17.2	57.3	4563	14	US-10-173-695-311	Sequence 311, App
c 300	17.2	57.3	609	17	US-10-027-632-102153	Sequence 102153, A	c 373	17.2	57.3	4563	14	US-10-173-697-311	Sequence 311, App
c 301	17.2	57.3	614	13	US-10-027-632-103268	Sequence 103268, A	c 374	17.2	57.3	4563	14	US-10-173-705-311	Sequence 311, App
c 302	17.2	57.3	614	17	US-10-027-632-103268	Sequence 103268, A	c 375	17.2	57.3	4563	14	US-10-174-576-311	Sequence 311, App
c 303	17.2	57.3	704	13	US-10-027-632-15994	Sequence 15994, A	c 376	17.2	57.3	4563	14	US-10-174-585-311	Sequence 311, App

377	17.2	57.3	4563	14	US-10-174-586-311	Sequence 311, App	450	17.2	57.3	4563	14	US-10-187-757-311	Sequence 311, App
378	17.2	57.3	4563	14	US-10-175-747-311	Sequence 311, App	451	17.2	57.3	4563	14	US-10-187-884-311	Sequence 311, App
379	17.2	57.3	4563	14	US-10-176-481-311	Sequence 311, App	452	17.2	57.3	4563	14	US-10-188-767-311	Sequence 311, App
380	17.2	57.3	4563	14	US-10-176-485-311	Sequence 311, App	453	17.2	57.3	4563	14	US-10-188-769-311	Sequence 311, App
381	17.2	57.3	4563	14	US-10-176-487-311	Sequence 311, App	454	17.2	57.3	4563	14	US-10-188-770-311	Sequence 311, App
382	17.2	57.3	4563	14	US-10-176-493-311	Sequence 311, App	455	17.2	57.3	4563	14	US-10-188-773-311	Sequence 311, App
383	17.2	57.3	4563	14	US-10-176-756-311	Sequence 311, App	456	17.2	57.3	4563	14	US-10-188-781-311	Sequence 311, App
384	17.2	57.3	4563	14	US-10-176-911-311	Sequence 311, App	457	17.2	57.3	4563	14	US-10-194-361-311	Sequence 311, App
385	17.2	57.3	4563	14	US-10-176-919-311	Sequence 311, App	458	17.2	57.3	4563	14	US-10-194-423-311	Sequence 311, App
386	17.2	57.3	4563	14	US-10-176-925-311	Sequence 311, App	459	17.2	57.3	4563	14	US-10-195-897-311	Sequence 311, App
387	17.2	57.3	4563	14	US-10-176-978-311	Sequence 311, App	460	17.2	57.3	4563	14	US-10-195-901-311	Sequence 311, App
388	17.2	57.3	4563	14	US-10-179-510-311	Sequence 311, App	461	17.2	57.3	4563	14	US-10-195-902-311	Sequence 311, App
389	17.2	57.3	4563	14	US-10-180-543-311	Sequence 311, App	462	17.2	57.3	4563	14	US-10-196-743-311	Sequence 311, App
390	17.2	57.3	4563	14	US-10-180-546-311	Sequence 311, App	463	17.2	57.3	4563	14	US-10-196-708-311	Sequence 311, App
391	17.2	57.3	4563	14	US-10-180-547-311	Sequence 311, App	464	17.2	57.3	4563	14	US-10-173-708-311	Sequence 311, App
392	17.2	57.3	4563	14	US-10-180-547-311	Sequence 311, App	465	17.2	57.3	4563	14	US-10-176-479-311	Sequence 311, App
393	17.2	57.3	4563	14	US-10-180-549-311	Sequence 311, App	466	17.2	57.3	4563	14	US-10-176-748-311	Sequence 311, App
394	17.2	57.3	4563	14	US-10-180-555-311	Sequence 311, App	467	17.2	57.3	4563	14	US-10-176-916-311	Sequence 311, App
395	17.2	57.3	4563	14	US-10-180-559-311	Sequence 311, App	468	17.2	57.3	4563	14	US-10-179-507-311	Sequence 311, App
396	17.2	57.3	4563	14	US-10-181-000-311	Sequence 311, App	469	17.2	57.3	4563	14	US-10-179-516-311	Sequence 311, App
397	17.2	57.3	4563	14	US-10-183-010-311	Sequence 311, App	470	17.2	57.3	4563	14	US-10-179-519-311	Sequence 311, App
398	17.2	57.3	4563	14	US-10-183-012-311	Sequence 311, App	471	17.2	57.3	4563	14	US-10-179-525-311	Sequence 311, App
399	17.2	57.3	4563	14	US-10-184-614-311	Sequence 311, App	472	17.2	57.3	4563	14	US-10-180-540-311	Sequence 311, App
400	17.2	57.3	4563	14	US-10-184-623-311	Sequence 311, App	473	17.2	57.3	4563	14	US-10-180-545-311	Sequence 311, App
401	17.2	57.3	4563	14	US-10-184-635-311	Sequence 311, App	474	17.2	57.3	4563	14	US-10-183-006-311	Sequence 311, App
402	17.2	57.3	4563	14	US-10-184-637-311	Sequence 311, App	475	17.2	57.3	4563	14	US-10-183-008-311	Sequence 311, App
403	17.2	57.3	4563	14	US-10-184-646-311	Sequence 311, App	476	17.2	57.3	4563	14	US-10-183-017-311	Sequence 311, App
404	17.2	57.3	4563	14	US-10-184-647-311	Sequence 311, App	477	17.2	57.3	4563	14	US-10-183-019-311	Sequence 311, App
405	17.2	57.3	4563	14	US-10-184-652-311	Sequence 311, App	478	17.2	57.3	4563	14	US-10-184-618-311	Sequence 311, App
406	17.2	57.3	4563	14	US-10-187-594-311	Sequence 311, App	479	17.2	57.3	4563	14	US-10-184-625-311	Sequence 311, App
407	17.2	57.3	4563	14	US-10-187-596-311	Sequence 311, App	480	17.2	57.3	4563	14	US-10-184-626-311	Sequence 311, App
408	17.2	57.3	4563	14	US-10-187-745-311	Sequence 311, App	481	17.2	57.3	4563	14	US-10-184-627-311	Sequence 311, App
409	17.2	57.3	4563	14	US-10-187-885-311	Sequence 311, App	482	17.2	57.3	4563	14	US-10-184-645-311	Sequence 311, App
410	17.2	57.3	4563	14	US-10-187-886-311	Sequence 311, App	483	17.2	57.3	4563	14	US-10-184-654-311	Sequence 311, App
411	17.2	57.3	4563	14	US-10-189-464-311	Sequence 311, App	484	17.2	57.3	4563	14	US-10-184-655-311	Sequence 311, App
412	17.2	57.3	4563	14	US-10-196-756-311	Sequence 311, App	485	17.2	57.3	4563	14	US-10-188-774-311	Sequence 311, App
413	17.2	57.3	4563	14	US-10-176-751-311	Sequence 311, App	486	17.2	57.3	4563	14	US-10-188-775-311	Sequence 311, App
414	17.2	57.3	4563	14	US-10-176-760-311	Sequence 311, App	487	17.2	57.3	4563	14	US-10-194-462-311	Sequence 311, App
415	17.2	57.3	4563	14	US-10-176-980-311	Sequence 311, App	488	17.2	57.3	4563	14	US-10-196-745-311	Sequence 311, App
416	17.2	57.3	4563	14	US-10-180-541-311	Sequence 311, App	489	17.2	57.3	4563	14	US-10-196-762-311	Sequence 311, App
417	17.2	57.3	4563	14	US-10-180-542-311	Sequence 311, App	490	17.2	57.3	4563	14	US-10-197-695-311	Sequence 311, App
418	17.2	57.3	4563	14	US-10-180-548-311	Sequence 311, App	491	17.2	57.3	4563	14	US-10-195-894-311	Sequence 311, App
419	17.2	57.3	4563	14	US-10-180-551-311	Sequence 311, App	492	17.2	57.3	4563	14	US-10-006-856A-259	Sequence 259, App
420	17.2	57.3	4563	14	US-10-180-998-311	Sequence 311, App	493	17.2	57.3	4563	14	US-10-176-484-311	Sequence 311, App
421	17.2	57.3	4563	14	US-10-180-999-311	Sequence 311, App	494	17.2	57.3	4563	14	US-10-176-753-311	Sequence 311, App
422	17.2	57.3	4563	14	US-10-183-013-311	Sequence 311, App	495	17.2	57.3	4563	14	US-10-176-917-311	Sequence 311, App
423	17.2	57.3	4563	14	US-10-184-612-311	Sequence 311, App	496	17.2	57.3	4563	14	US-10-176-982-311	Sequence 311, App
424	17.2	57.3	4563	14	US-10-184-616-311	Sequence 311, App	497	17.2	57.3	4563	14	US-10-179-506-311	Sequence 311, App
425	17.2	57.3	4563	14	US-10-184-617-311	Sequence 311, App	498	17.2	57.3	4563	14	US-10-179-513-311	Sequence 311, App
426	17.2	57.3	4563	14	US-10-184-622-311	Sequence 311, App	499	17.2	57.3	4563	14	US-10-179-514-311	Sequence 311, App
427	17.2	57.3	4563	14	US-10-184-628-311	Sequence 311, App	500	17.2	57.3	4563	14	US-10-179-522-311	Sequence 311, App
428	17.2	57.3	4563	14	US-10-184-629-311	Sequence 311, App							
429	17.2	57.3	4563	14	US-10-184-630-311	Sequence 311, App							
430	17.2	57.3	4563	14	US-10-184-631-311	Sequence 311, App							
431	17.2	57.3	4563	14	US-10-184-632-311	Sequence 311, App							
432	17.2	57.3	4563	14	US-10-184-636-311	Sequence 311, App							
433	17.2	57.3	4563	14	US-10-184-640-311	Sequence 311, App							
434	17.2	57.3	4563	14	US-10-184-650-311	Sequence 311, App							
435	17.2	57.3	4563	14	US-10-184-651-311	Sequence 311, App							
436	17.2	57.3	4563	14	US-10-187-588-311	Sequence 311, App							
437	17.2	57.3	4563	14	US-10-187-597-311	Sequence 311, App							
438	17.2	57.3	4563	14	US-10-187-598-311	Sequence 311, App							
439	17.2	57.3	4563	14	US-10-187-600-311	Sequence 311, App							
440	17.2	57.3	4563	14	US-10-187-601-311	Sequence 311, App							
441	17.2	57.3	4563	14	US-10-187-602-311	Sequence 311, App							
442	17.2	57.3	4563	14	US-10-187-603-311	Sequence 311, App							
443	17.2	57.3	4563	14	US-10-187-741-311	Sequence 311, App							
444	17.2	57.3	4563	14	US-10-187-743-311	Sequence 311, App							
445	17.2	57.3	4563	14	US-10-187-746-311	Sequence 311, App							
446	17.2	57.3	4563	14	US-10-187-747-311	Sequence 311, App							
447	17.2	57.3	4563	14	US-10-187-751-311	Sequence 311, App							
448	17.2	57.3	4563	14	US-10-187-753-311	Sequence 311, App							
449	17.2	57.3	4563	14	US-10-187-754-311	Sequence 311, App							

ALIGNMENTS

RESULT 1

US-10-010-476-9/c

; Sequence 9, Application US/10010476

; Publication No. US20030114402A1

; GENERAL INFORMATION:

; APPLICANT: Reich, No. US20030114402A1bert O.

; FLYNN, James

; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5

; CORRESPONDENCE ADDRESS: METHYLTRANSFERASE AND METHODS FOR USE THEREOF

; NUMBER OF SEQUENCES: 110

; ADDRESS: Merchant, Gould, Smith, Edell, Welter & Schmidt

; STREET: 11150 Santa Monica Boulevard, Suite 400

; CITY: Los Angeles

; STATE: CA

; COUNTRY: USA

; ZIP: 90025

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-010-476-9

Query Match 100.0%; Score 30; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30
DB 30 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 1

RESULT 2

US-10-010-476-10
Sequence 10, Application US/10010476
Publication No. US20030114402A1
GENERAL INFORMATION:
APPLICANT: Reich, No. US20030114402Albert O.
Flynn, James
TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
METHYLTRANSFERASE AND METHODS FOR USE THEREOF
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S

REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-010-476-10

Query Match 100.0%; Score 30; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30
DB 1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30

RESULT 3

US-10-010-476-13
Sequence 13, Application US/10010476
Publication No. US20030114402A1
GENERAL INFORMATION:
APPLICANT: Reich, No. US20030114402Albert O.
Flynn, James
TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
METHYLTRANSFERASE AND METHODS FOR USE THEREOF
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-010-476-13

Query Match

100.0%; Score 30; DB 15; Length 50;

```
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 30
Db 11 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 40

RESULT 4
US-10-010-476-14
; Sequence 14, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402A1bert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-010-476-14
Query Match 73.3%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATCCTTGGCCCGCCCTTGAAT 26
Db 1 ATCCTTGGCCCGCCCTTGAAT 22

RESULT 5
US-10-664-705-74
; Sequence 74, Application US/10664705
; Publication No. US20040152107A1
; GENERAL INFORMATION:
; APPLICANT: Altar, Anthony C.
; APPLICANT: Laeng, Pascal
; APPLICANT: Young, Theresa A.

Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 30
Db 11 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 40

RESULT 6
US-10-425-115-93962
; Sequence 93962, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 93962
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_185680C.1
US-10-425-115-93962
Query Match 70.0%; Score 21; DB 18; Length 593;
Best Local Similarity 82.8%; Pred. No. 29;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 29
Db 403 CTGGATCCTTGCCTCCCGCCCTTGAATCCC 431

RESULT 7
US-10-116-802-513/C
; Sequence 513, Application US/10116802
; Publication No. US20030065157A1
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
```



```

Query Match          66.7%; Score 20; DB 17; Length 1760;
Best Local Similarity 82.1%; Pred. No. 68;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      2  TGGATCCTTGGCCCGCCCTTGAATTCC 29
      ||| ||| ||| ||| ||| ||| |||
Db      431 TGGGTGCTCGCCCGCCCACTCGAATTCC 404

RESULT 11
US-09-783-590-3630/c
; Sequence 3630, Application US/09783590
; Patent No. US20020110850A1
GENERAL INFORMATION:

```

; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3630
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (231)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (318)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (343)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (349)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (354)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (387)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (405)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (409)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (415)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (448)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (452)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-3630

Query Match 66.0%; Score 19.8; DB 9; Length 456;
Best Local Similarity 84.0%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 TCCTTGGCCCGCCCTTGAATTC 30
DB 397 TCCTTGGCTCCNNCCCTTGAATTC 373

RESULT 12
US-10-291-265-476/c
; Sequence 476, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 476
; LENGTH: 4385
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-291-265-476

Query Match 65.3%; Score 19.6; DB 17; Length 4385;
Best Local Similarity 84.6%; Pred. No. 93;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATCCTTGCCCGCCCTTGAATTC 30
DB 94 ACCCTTGCCCAACCTTGAATTC 69

RESULT 13
US-10-364-505-9/c
; Sequence 9, Application US/10364505
; Publication No. US20030219787A1
; GENERAL INFORMATION:
; APPLICANT: Kere, Juha
; APPLICANT: Taipale, Mikko
; APPLICANT: No. US20030219787A1ola-Hemmi, Jaana
; APPLICANT: Kaminen, Nina
; TITLE OF INVENTION: NOVEL HUMAN GENE FUNCTIONALLY RELATED TO DYSLEXIA
; FILE REFERENCE: 0933-0199P
; CURRENT APPLICATION NUMBER: US/10/364,505
; CURRENT FILING DATE: 2003-02-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human DYXC1 chromosomal gene region, nucleotides
; OTHER INFORMATION: 150001-200000
US-10-364-505-9

Query Match 65.3%; Score 19.6; DB 17; Length 50000;
Best Local Similarity 84.6%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATCCTTGCCCGCCCTTGAATTC 30
DB 42561 ACCCTTGCCCAACCTTGAATTC 42536

RESULT 14
US-10-681-199-9/c
; Sequence 9, Application US/10681199
; Publication No. US20040138441A1

```
; GENERAL INFORMATION:
; APPLICANT: KERE, Juha
; TITLE OF INVENTION: NOVEL HUMAN GENE FUNCTIONALLY RELATED TO DYSLEXIA
; FILE REFERENCE: 0933-0214P
; CURRENT APPLICATION NUMBER: US/10/681,199
; CURRENT FILING DATE: 2003-10-09
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: human DYXC1 chromosomal gene region, nucleotides
; OTHER INFORMATION: 150001-200000
US-10-681-199--9

Query Match      65.3%; Score 19.6; DB 18; Length 50000;
Best Local Similarity 84.6%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      5 ATCTTGGCCCGCCCTTGAATTC 30
Db      42561 ACCCTTGGCCCAACCTTGAATTC 42536

RESULT 15
US-10-719-993-6802
; Sequence 6802, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6802
; LENGTH: 191597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-6802

Query Match      65.3%; Score 19.6; DB 18; Length 191597;
Best Local Similarity 84.6%; Pred. No. 68;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CTGGATCCTTGGCCCGCCCTTGAAT 26
Db      144967 CTGAATCCTTGTCTTGGCCCTTGAAT 144992

RESULT 16
US-10-430-201-2349
; Sequence 2349, Application US/10430201
; Publication No. US20040162679A1
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment
; FILE REFERENCE: 40716 (IP-010)
; CURRENT APPLICATION NUMBER: US/10/430,201
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/370,114
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 4879
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2349
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (384)..(399)
; OTHER INFORMATION: n is a, c, g, or t
US-10-430-201-2349

Query Match      64.7%; Score 19.4; DB 18; Length 540;
Best Local Similarity 79.3%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 TGGATCCTTGGCCCGCCCTTGAATTC 30
Db      276 TGTCTCTTCCCTGCCCCCTTGAATCCAC 304

RESULT 17
US-10-430-201-2350
; Sequence 2350, Application US/10430201
; Publication No. US20040162679A1
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment
; FILE REFERENCE: 40716 (IP-010)
; CURRENT APPLICATION NUMBER: US/10/430,201
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/370,114
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 4879
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2350
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (384)..(399)
; OTHER INFORMATION: n is a, c, g, or t
US-10-430-201-2350

Query Match      64.7%; Score 19.4; DB 18; Length 540;
Best Local Similarity 79.3%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 TGGATCCTTGGCCCGCCCTTGAATTC 30
Db      276 TGTCTCTTCCCTGCCCCCTTGAATCCAC 304

RESULT 18
US-10-369-493-25548/c
; Sequence 25548, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25548
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25548

Query Match      64.7%; Score 19.4; DB 17; Length 1041;
Best Local Similarity 79.3%; Pred. No. 1.3e+02;
```

```
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TGGATCCTTGGCCCGCCCTTGAATTCCC 30
Db 1030 TGGATCCTTGGCCCTTCTCTTGAGTCTC 1002

RESULT 19
US-09-997-722-151/c
; Sequence 151, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 151
; LENGTH: 96593
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(312)
; OTHER INFORMATION: "n" at positions 1 through 312 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3386)..(4940)
; OTHER INFORMATION: "n" at positions 3386 through 4940 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12981)..(13197)
; OTHER INFORMATION: "n" at positions 12981 through 13197 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65121)..(65140)
; OTHER INFORMATION: "n" at positions 65121 through 65140 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (82761)..(82780)
; OTHER INFORMATION: "n" at positions 82761 through 82780 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (90594)..(90613)
; OTHER INFORMATION: "n" at positions 90594 through 90613 can be any base.
US-09-997-722-151

Query Match 64.7%; Score 19.4; DB 11; Length 96593;
Best Local Similarity 79.3%; Pred. No. 87;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCC 29
Db 9903 CTGGCTCTGGCCCGCCCTAGAGCTGC 9875

RESULT 20
US-10-723-860-7320/c
; Sequence 7320, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlocnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods of Screening for Soft Tissue Sarcoma Modulators
```

```
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7320
; LENGTH: 1189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-7320

Query Match 64.0%; Score 19.2; DB 18; Length 1189;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 7 CTTTGCCCGCCCTTGAATTCCTCC 30
Db 104 CTTGACCCCTGCGCCCTTGAATTCCTCC 81

RESULT 21
US-10-425-115-97766/c
; Sequence 97766, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 97766
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_20673C.1
US-10-425-115-97766

Query Match 63.3%; Score 19; DB 18; Length 371;
Best Local Similarity 81.5%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATT 27
Db 233 CTTGACCCCTGACCCCGCCCTTGAATT 207

RESULT 22
US-10-963-24928/c
; Sequence 24928, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

SEQ ID NO 24928

LENGTH: 741

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT_MRT4530_29866C.1

US-10-437-963-24928

Query Match 63.3%; Score 19; DB 18; Length 741;

Best Local Similarity 81.5%; Pred. No. 1.9e+02;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GGATCCTTGGCCCGCCCTTGAATCCC 29

Db 342 GGATCCTGTGCTCTCCCTTGAATCCC 316

RESULT 23

US-10-010-476-11/c

Sequence 11, Application US/10010476

Publication No. US20030114402A1

GENERAL INFORMATION:

APPLICANT: Reich, No. US20030114402Albert O.

APPLICANT: Flynn, James

TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5

NUMBER OF SEQUENCES: 110

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt

STREET: 11150 Santa Monica Boulevard, Suite 400

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/010,476

FILING DATE: 07-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/485,071

FILING DATE: 12-Jun-1998

APPLICATION NUMBER: 60/057,411

FILING DATE: 29-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Canady, Karen S

REGISTRATION NUMBER: 39,927

REFERENCE/DOCKET NUMBER: 30794.30WO01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 310 445-1140

TELEFAX: 310 445-9031

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-10-010-476-11

Query Match

Best Local Similarity 62.7%; Score 18.8; DB 15; Length 30;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30

Db 30 CTGGATCCTTTTGACGTCAATTTGAATCCC 1

RESULT 24

US-10-010-476-12

Sequence 12, Application US/10010476

Publication No. US20030114402A1

GENERAL INFORMATION:

APPLICANT: Reich, No. US20030114402Albert O.

APPLICANT: Flynn, James

TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5

NUMBER OF SEQUENCES: 110

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt

STREET: 11150 Santa Monica Boulevard, Suite 400

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/010,476

FILING DATE: 07-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/485,071

FILING DATE: 12-Jun-1998

APPLICATION NUMBER: 60/057,411

FILING DATE: 29-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Canady, Karen S

REGISTRATION NUMBER: 39,927

REFERENCE/DOCKET NUMBER: 30794.30WO01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 310 445-1140

TELEFAX: 310 445-9031

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-10-010-476-12

Query Match

Best Local Similarity 62.7%; Score 18.8; DB 15; Length 30;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30

Db 1 CTGGATCCTTTTGACGTCAATTTGAATCCC 30

RESULT 25

US-10-029-386-13283

Sequence 13283, Application US/10029386

Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE

FILE REFERENCE: AROMICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
; SEQ ID NO 13283
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR14.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.46
; OTHER INFORMATION: SWISSPROT HIT: Q35799, EVALUAE 4.20e+00
; OTHER INFORMATION: EST HUMAN HIT: BI825694.1, EVALUAE 3.00e-80
; OTHER INFORMATION: NT HIT: gill4751632, EVALUAE 2.00e-80
US-10-029-386-13283

Query Match          62.7%; Score 18.8; DB 16; Length 577;
Best Local Similarity 76.7%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 30
Db 170 CTGTGCTTGGCCACCATCTGTATTC 199

RESULT 26
US-10-027-632-28086/c
; Sequence 28086, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28086
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-28086

Query Match          62.7%; Score 18.8; DB 13; Length 680;
Best Local Similarity 76.7%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 30
Db 53 CTGGCTGCTTCCCTGCGCGCTGAATGCC 24

RESULT 27
US-10-027-632-28086/c
; Sequence 28086, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28086
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-28086

Query Match          62.7%; Score 18.8; DB 13; Length 680;
Best Local Similarity 76.7%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 30
Db 53 CTGGCTGCTTCCCTGCGCGCTGAATGCC 24

RESULT 28
US-10-198-846-596
; Sequence 596, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 596
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 8, 10, 15, 18, 20, 22, 25, 27, 33, 35, 41, 43, 46, 47,
; LOCATION: 50, 56, 57, 59, 62, 65, 66, 68, 70, 78, 84, 88, 98, 109,
; LOCATION: 115, 116, 118, 120, 122, 124, 125, 127, 129, 130, 171, 308,
; LOCATION: 390, 395, 438, 452, 462, 468, 473, 475, 482, 488, 491, 575
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 578, 583, 595, 614, 669, 700
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-596

Query Match          62.7%; Score 18.8; DB 14; Length 708;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 CTGGCCCGCCCTTGAATTC 30
```

```
; SEQ ID NO 167839
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167839

Query Match      62.7%; Score 18.8; DB 17; Length 816;
Best Local Similarity 76.7%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 167839
LENGTH: 816
TYPE: DNA
ORGANISM: Human
US-10-027-632-167839

Query Match      62.7%; Score 18.8; DB 13; Length 816;
Best Local Similarity 76.7%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 167839
LENGTH: 816
TYPE: DNA
ORGANISM: Human
US-10-027-632-167839

Query Match      62.7%; Score 18.8; DB 11; Length 1051;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 167839
LENGTH: 816
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-047-449
```

Query Match 62.7%; Score 18.8; DB 17; Length 2021;
Best Local Similarity 76.7%; Pred. No. 2.2e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
| | | | | | | | | | | | | | | | | | | | | |
Db 101 CGGGCTCCTGCGCCGCGCTTCTCTCCC 72

RESULT 33

US-10-094-749-137

; Sequence 137, Application US/10094749

; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHIKO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOTYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; PRIOR FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 137

; LENGTH: 2048

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-094-749-137

Query Match 62.7%; Score 18.8; DB 17; Length 2048;
Best Local Similarity 76.7%; Pred. No. 2.2e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
| | | | | | | | | | | | | | | | | | | | | |
Db 617 CTGGCTCCTTCCCTGCGCGCTGAATGCC 646

RESULT 34

US-10-335-053-162

; Sequence 162, Application US/10335053

; Publication No. US20040241653A1

; GENERAL INFORMATION:

; APPLICANT: Quark Biotech, Inc.

; TITLE OF INVENTION: Methods for identifying marker genes for cancer

; FILE REFERENCE: 68733-A; 070/US1

; CURRENT APPLICATION NUMBER: US/10/335,053

; CURRENT FILING DATE: 2003-03-27

; PRIOR APPLICATION NUMBER: 60/345,317

; PRIOR FILING DATE: 2001-12-31

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 162

; LENGTH: 2524

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-335-053-162

Query Match 62.7%; Score 18.8; DB 18; Length 2524;
Best Local Similarity 76.7%; Pred. No. 2.1e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
| | | | | | | | | | | | | | | | | | | | | |
Db 1091 CTGGCTCCTTCCCTGCGCGCTGAATGCC 1120

RESULT 35

US-10-437-963-77151

; Sequence 77151, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 77151

; LENGTH: 2913

; TYPE: DNA

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_77077C.1

US-10-437-963-77151

Query Match 62.7%; Score 18.8; DB 18; Length 2913;
Best Local Similarity 76.7%; Pred. No. 2.1e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
| | | | | | | | | | | | | | | | | | | | | |
Db 1035 CGGGCTCCTTGGCCCGCGCTGGACTACGC 1064

RESULT 36

US-09-764-891-9763

; Sequence 9763, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9763

; LENGTH: 9370

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-891-9763

Query Match 62.7%; Score 18.8; DB 10; Length 9370;
Best Local Similarity 76.7%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;


```

RESULT 46
US-10-425-115-86331/c
; Sequence 86331, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/425.115

```

```

RESULT 48
US-10-027-632-7978
; Sequence 7978, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIORITY APPLICATION NUMBER: US 60/218,006

```

```
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7978
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-7978
```

```
Query Match          62.0%; Score 18.6; DB 13; Length 646;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      6 TCCTTGCCCGCCCTTGAAATCCC 30
          |||||
Db      220 TTCCTGCCCGCCCATTTGAATCCC 244
```

RESULT 49

```
US-10-027-632-7977
; Sequence 7977, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7977
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-7977
```

```
Query Match          62.0%; Score 18.6; DB 17; Length 646;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      6 TCCTTGCCCGCCCTTGAAATCCC 30
          |||||
Db      220 TTCCTGCCCGCCCATTTGAATCCC 244
```

RESULT 50

```
US-10-027-632-7978
; Sequence 7978, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7978
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-7978
```

```
Query Match          62.0%; Score 18.6; DB 17; Length 646;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      6 TCCTTGCCCGCCCTTGAAATCCC 30
          |||||
Db      220 TTCCTGCCCGCCCATTTGAATCCC 244
```

```
Search completed: April 25, 2005, 15:33:50
Job time : 394.483 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:07:14 ; Search time 1055.17 Seconds

(without alignments)
2296.080 Million cell updates/sec

Title: US-10-010-476-13

Perfect score: 50

Sequence: 1 CTTACCCACCTGGATCCTT.....TTGAATTCACACCTCCAC 50

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ste.*

12: gb_sy.*

13: gb_uni.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	50	6	BD078129 Modulator
2	30	60.0	30	6	BD078125 Modulator
3	30	60.0	30	6	BD078126 Modulator
4	26.6	53.2	10805	1	AE005937 Caulobact
5	26.6	53.2	141056	9	AL354836 Human DNA
6	26.6	53.2	152806	2	AC129725 Rattus no
7	26.6	53.2	194785	2	AC136047 Rattus no
8	26.6	53.2	239910	2	AC105637 Rattus no
9	26.6	53.2	241243	2	AC094206 Rattus no
10	26.4	52.8	22444	6	AX695482 Sequence
11	26.4	52.8	217726	10	AL928680 Mouse DNA
12	26	52.0	98501	9	AP005019 Homo sapi
13	26	52.0	134729	10	AC128662 Mus muscu
14	26	52.0	251756	10	AC119801 Mus muscu
15	25.6	51.2	1032	4	AB115779 Bos tauru
16	25.6	51.2	204144	2	AB117870 Rattus no
17	25.6	51.2	204513	2	AC129955 Bos tauru
18	25.6	51.2	209857	2	AC096053 Rattus no
19	25.6	51.2	211988	4	AC129959 Bos tauru

C	20	25.4	50.8	98204	5	AL645800	Zebrafish
C	21	25.4	50.8	132033	9	AC099796	Homo sapi
C	22	25.4	50.8	166549	2	AC025990	Homo sapi
C	23	25.4	50.8	201933	2	AC141729	Apis mell
C	24	25.2	50.4	422	6	CQ135284	Sequence
C	25	25.2	50.4	422	6	CQ148287	Sequence
C	26	25.2	50.4	705	6	AX525599	Sequence
C	27	25.2	50.4	4562	10	D88672	Rattus norv
C	28	25.2	50.4	127035	2	AC149255	Papio anu
C	29	25.2	50.4	148133	9	CNS00000T	Human chr
C	30	25.2	50.4	155358	9	AL354858	Human DNA
C	31	25.2	50.4	169164	2	AC134747	Rattus no
C	32	25.2	50.4	188447	2	AC095795	Rattus no
C	33	25.2	50.4	190177	2	AC009959	Homo sapi
C	34	25.2	50.4	204020	10	AL929424	Mouse DNA
C	35	25.2	50.4	209383	10	AC121604	Mus muscu
C	36	25.2	50.4	210493	2	AC146886	Callithri
C	37	25.2	50.4	216177	2	AC146883	Callithri
C	38	25.2	50.4	218761	2	AX294436	Mus muscu
C	39	25.2	50.4	228016	2	AC146674	Callithri
C	40	25.2	50.4	232953	2	AC140217	Mus muscu
C	41	25.2	50.4	237359	2	AC127753	Rattus no
C	42	25.2	50.4	263882	2	AC107344	Rattus no
C	43	25	50.0	53950	2	AC015310	Drosophil
C	44	25	50.0	101445	9	AL133411	Human DNA
C	45	25	50.0	131007	9	HS998C11	Human DNA
C	46	25	50.0	131769	9	AC104506	Homo sapi
C	47	25	50.0	152858	9	AC093428	Homo sapi
C	48	25	50.0	188766	3	AC010708	Drosophil
C	49	25	50.0	191877	9	AC007207	Homo sapi
C	50	25	50.0	223162	2	AC137176	Rattus no
C	51	25	50.0	242875	2	AC096494	Rattus no
C	52	25	50.0	243428	6	AX923202	Sequence
C	53	25	50.0	244594	2	AC112470	Rattus no
C	54	25	50.0	260981	2	AC095001	Rattus no
C	55	25	50.0	262091	2	AC098335	Rattus no
C	56	25	50.0	264466	2	AC126956	Rattus no
C	57	25	50.0	300619	3	AE003508	Drosophil
C	58	24.8	49.6	177836	2	AC149559	Papio anu
C	59	24.8	49.6	192525	2	AC109540	Rattus no
C	60	24.8	49.6	216000	2	AL731715	Mus muscu
C	61	24.8	49.6	232225	10	AL662876	Mouse DNA
C	62	24.6	49.2	154844	9	AC017000	Homo sapi
C	63	24.6	49.2	169005	9	AC034198	Homo sapi
C	64	24.6	49.2	184716	9	AC018836	Homo sapi
C	65	24.6	49.2	187624	9	AC093768	Homo sapi
C	66	24.6	49.2	208558	2	AC128403	Rattus no
C	67	24.6	49.2	211012	9	AC018843	Homo sapi
C	68	24.6	49.2	248338	2	AC106101	Rattus no
C	69	24.6	49.2	309430	2	AC012209	Homo sapi
C	70	24.6	49.2	347800	1	AP000060	Aeropyrum
C	71	24.4	48.8	1637	9	HS011714	Homo sapi
C	72	24.4	48.8	3197	9	BC040529	Homo sapi
C	73	24.4	48.8	6960	6	AX347300	Sequence
C	74	24.4	48.8	21319	9	HUMABERY	Human anion
C	75	24.4	48.8	22843	2	AC015232	Drosophil
C	76	24.4	48.8	91061	2	AC087139	Mus muscu
C	77	24.4	48.8	101839	9	AL445486	Human DNA
C	78	24.4	48.8	108752	10	AF084363	Mus muscu
C	79	24.4	48.8	139488	9	AC003043	Homo sapi
C	80	24.4	48.8	152443	3	AC009208	Drosophil
C	81	24.4	48.8	156202	2	AC148840	Pongo pyg
C	82	24.4	48.8	160267	2	AC101862	Mus muscu
C	83	24.4	48.8	166664	9	AL353573	Human DNA
C	84	24.4	48.8	168961	2	AC130786	Papio anu
C	85	24.4	48.8	176647	2	AC130188	Papio anu
C	86	24.4	48.8	177076	9	AP000487	Homo sapi
C	87	24.4	48.8	179325	3	AC007418	Drosophil
C	88	24.4	48.8	180335	3	AC146276	Pan trogl
C	89	24.4	48.8	187110	10	AC131303	Mus muscu
C	90	24.4	48.8	189700	2	AC123381	Rattus no
C	91	24.4	48.8	194719	2	AC026173	Homo sapi
C	92	24.4	48.8	194799	9	AC006430	Homo sapi

239	23.4	46.8	110000	2	AC101867.1	Continuation (2 of
240	23.4	46.8	119258	10	AL731860	Mouse DNA
241	23.4	46.8	120249	2	AP000858	Homo sapi
242	23.4	46.8	127295	9	AP002346	Homo sapi
243	23.4	46.8	132636	2	AC108369	Pan trogl
244	23.4	46.8	132764	9	AC073842	Homo sapi
245	23.4	46.8	137862	2	AL356780	Homo sapi
246	23.4	46.8	138969	8	OSJN00075	Mouse DNA
247	23.4	46.8	143406	2	BX465195	Oryza sat
248	23.4	46.8	143481	9	AC013472	Homo sapi
249	23.4	46.8	152376	2	AC116682	Mus muscu
250	23.4	46.8	162921	2	AC013528	Homo sapi
251	23.4	46.8	167854	2	AC092483	Homo sapi
252	23.4	46.8	169638	9	AC009433	Homo sapi
253	23.4	46.8	171061	10	AC104833	Mus Muscu
254	23.4	46.8	178862	9	AC011847	Homo sapi
255	23.4	46.8	179104	9	AC063963	Homo sapi
256	23.4	46.8	179155	9	AL353748	Human DNA
257	23.4	46.8	179221	9	AC114493	Homo sapi
258	23.4	46.8	185041	2	AC128276	Rattus no
259	23.4	46.8	185854	2	AC027736	Homo sapi
260	23.4	46.8	186134	10	AL591390	Mouse DNA
261	23.4	46.8	187979	10	AC123854	Mus muscu
262	23.4	46.8	190644	10	AC128272	Mus muscu
263	23.4	46.8	205816	2	AC145411	Boe tauru
264	23.4	46.8	206000	10	AC102248	Mus muscu
265	23.4	46.8	209706	2	AC119611	Rattus no
266	23.4	46.8	214019	2	AC064803	Mus muscu
267	23.4	46.8	214733	2	AC135154	Rattus no
268	23.4	46.8	221642	2	AC115978	Mus muscu
269	23.4	46.8	222186	2	AC105596	Rattus no
270	23.4	46.8	223020	9	AL162424	Human DNA
271	23.4	46.8	225216	10	AL844888	Mouse DNA
272	23.4	46.8	227341	2	AC127771	Rattus no
273	23.4	46.8	227781	2	AC132915	Mus muscu
274	23.4	46.8	228212	2	AC114168	Rattus no
275	23.4	46.8	232993	2	AC098761	Rattus no
276	23.4	46.8	240992	2	AC131467	Rattus no
277	23.4	46.8	242413	2	AC106264	Rattus no
278	23.4	46.8	247575	2	AC133047	Rattus no
279	23.4	46.8	264754	2	AC128542	Rattus no
280	23.4	46.8	277621	2	AC095191	Rattus no
281	23.4	46.8	286524	2	AC105802	Rattus no
282	23.2	46.4	386	8	HVU234795	Hordeum v
283	23.2	46.4	2560	10	BC034106	Mus muscu
284	23.2	46.4	4558	10	AK172968	Human DNA
285	23.2	46.4	2839	9	HSL33C6	Continuation (4 of
286	23.2	46.4	68399	2	AC115459.3	Homo sapi
287	23.2	46.4	68646	2	AC105040	Homo sapi
288	23.2	46.4	89107	9	AC026695	Homo sapi
289	23.2	46.4	121074	2	CR456681	Danio rer
290	23.2	46.4	139919	9	AC002433	Homo sapi
291	23.2	46.4	161682	2	AC024438	Homo sapi
292	23.2	46.4	169565	2	AC079108	Homo sapi
293	23.2	46.4	173547	2	AC073637	Homo sapi
294	23.2	46.4	178338	9	AL158068	Human DNA
295	23.2	46.4	182010	2	AC118533	Rattus no
296	23.2	46.4	188903	2	AC145524	Papio ham
297	23.2	46.4	190952	2	AC119020	Rattus no
298	23.2	46.4	198537	9	AC068790	Homo sapi
299	23.2	46.4	207424	2	AC128853	Rattus no
300	23.2	46.4	221160	2	AC142129	Rattus no
301	23.2	46.4	235150	2	AC073776	Mus muscu
302	23.2	46.4	235011	2	AC105523	Rattus no
303	23.2	46.4	239456	2	AC108995	Rattus no
304	23.2	46.4	240216	2	AC107096	Rattus no
305	23.2	46.4	242624	2	AC128443	Rattus no
306	23.2	46.4	249561	2	AC116095	Rattus no
307	23.2	46.4	270094	2	AC106083	Rattus no
308	23.2	46.4	270128	2	AC103000	Rattus no
309	23	46.0	197	6	C0749386	Sequence
310	23	46.0	201	11	BV204596	sgm21570
311	23	46.0	512	6	AR415018	Sequence
6	512	46.0	512	6	AX971852	Sequence
6	512	46.0	512	6	BD110571	EST and e
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6	512	46.0	1900	9	AK057490	Homo sapi
6						

385	22.8	45.6	2380	9	AF119248	AF119248 Homo sapi	458	22.8	45.6	158050	9	AC023394	Homo sapi
386	22.8	45.6	3022	10	AF543842	AF543842 Mesocric	459	22.8	45.6	158073	9	AC093756	Homo sapi
387	22.8	45.6	3243	10	AV423767	AV423767 Mesocric	c 460	22.8	45.6	159082	10	AC124408	Mus muscu
388	22.8	45.6	3353	10	AV423768	AV423768 Mesocric	461	22.8	45.6	159138	10	AC124403	Mus muscu
389	22.8	45.6	3406	10	AF543843	AF543843 Mesocric	c 462	22.8	45.6	159525	8	AC093493	Oryza sat
390	22.8	45.6	4405	9	F202550S03	F202552 Homo sapi	c 463	22.8	45.6	161179	2	AC066601	Homo sapi
391	22.8	45.6	4460	6	BD190949	BD190949 Human methy	464	22.8	45.6	162060	2	AC007515	Homo sapi
392	22.8	45.6	7292	9	HSCST3G	HS2255 Human CSTR	465	22.8	45.6	162516	9	AC130464	Homo sapi
393	22.8	45.6	8216	9	AP001617	AP001617 Homo sapi	466	22.8	45.6	164493	2	AC078825	Homo sapi
394	22.8	45.6	8280	6	CQ797966	CQ797966 Sequence	c 467	22.8	45.6	164899	2	AC084218	Oryza sat
395	22.8	45.6	8280	9	AF161221	AF161221 Homo sapi	468	22.8	45.6	166247	2	AC135876	Rattus no
396	22.8	45.6	23401	9	AC090061	AC090061 Homo sapi	469	22.8	45.6	166279	9	AC135876	Human chr
397	22.8	45.6	28993	3	LMFLS808	AL354572 Leishmani	c 470	22.8	45.6	167792	9	CNS01DRQ	AC026163 Homo chr
398	22.8	45.6	46534	9	AC010077	AL590243 Homo sapi	471	22.8	45.6	167998	9	AL353662	Human DNA
399	22.8	45.6	51700	9	AL590243	AL590243 Human DNA	472	22.8	45.6	168522	2	AC024302	Homo sapi
400	22.8	45.6	51842	9	EX005143	EX005143 Human DNA	473	22.8	45.6	168700	2	AC009115	Homo sapi
401	22.8	45.6	58454	2	AC100233	AC100233 Mus muscu	474	22.8	45.6	169357	2	AC102221	Mus muscu
402	22.8	45.6	64080	9	AL133217	AL133217 Human DNA	c 475	22.8	45.6	169948	8	AP003871	AC079183 Oryza sat
403	22.8	45.6	65016	2	AC090620	AC090620 Homo sapi	c 476	22.8	45.6	170233	10	AC079183	Mus muscu
404	22.8	45.6	65241	3	AC084458	AC084458 Caenorhab	c 477	22.8	45.6	170610	2	AC023935	Homo sapi
405	22.8	45.6	68770	9	AC006985	AC006985 Homo sapi	c 478	22.8	45.6	171126	2	AC138306	Mus muscu
406	22.8	45.6	69278	2	AC101543	AC101543 Mus muscu	c 479	22.8	45.6	171296	9	CNS05C86	Human chr
407	22.8	45.6	69346	2	AC091148	AC091148 Homo sapi	480	22.8	45.6	171994	2	AC147516	Rattus no
408	22.8	45.6	71590	9	CR388219	CR388219 Human DNA	c 481	22.8	45.6	172457	2	AC122562	Mus muscu
409	22.8	45.6	73397	2	AC101427	AC101427 Mus muscu	482	22.8	45.6	172533	2	AC102195	Mus muscu
410	22.8	45.6	80658	9	AC107373	AC107373 Homo sapi	c 483	22.8	45.6	173357	10	AC135362	Mus muscu
411	22.8	45.6	84015	2	CR759782	CR759782 Homo sapi	c 484	22.8	45.6	173910	2	AC147398	Sus scrofa
412	22.8	45.6	84793	5	AC096684	AC096684 Takifugu	485	22.8	45.6	174124	2	AC102872	Mus muscu
413	22.8	45.6	85368	2	CR759839	CR759839 Homo sapi	486	22.8	45.6	174724	2	AC140096	Pan trogl
414	22.8	45.6	87917	2	AC016261	AC016261 Homo sapi	c 487	22.8	45.6	175262	9	AC021491	Homo sapi
415	22.8	45.6	88459	9	AL662849	AL662849 Human DNA	c 488	22.8	45.6	175711	2	AC022608	Homo sapi
416	22.8	45.6	92693	9	AL157766	AL157766 Human DNA	489	22.8	45.6	176064	2	AC123915	Mus muscu
417	22.8	45.6	95521	2	AC134499	AC134499 Rattus no	c 490	22.8	45.6	176354	2	EX855600	Mus muscu
418	22.8	45.6	100000	9	AP000502	AP000502 Homo sapi	c 491	22.8	45.6	176855	2	AC149557	Papio anu
419	22.8	45.6	101762	5	AL928854	AL928854 Zebrafish	c 492	22.8	45.6	177255	2	AC123235	Rattus no
420	22.8	45.6	102314	2	AC142621	AC142621 Macaca mu	493	22.8	45.6	177730	9	AC087844	Homo sapi
421	22.8	45.6	102781	2	AF176680	AF176680 Homo sapi	494	22.8	45.6	178460	2	AL139040	Homo sapi
422	22.8	45.6	105736	2	AC090678	AC090678 Homo sapi	c 495	22.8	45.6	178755	9	AC090673	Homo sapi
423	22.8	45.6	105801	2	EX537159	EX537159 Homo sapi	c 496	22.8	45.6	178868	9	AL357752	Human DNA
424	22.8	45.6	106182	2	CR788238	CR788238 Homo sapi	c 497	22.8	45.6	179242	2	AC145539	Didelphis
425	22.8	45.6	109646	9	HSNHC78S22	AF019413 Homo sapi	498	22.8	45.6	179537	2	AC013513	Homo sapi
426	22.8	45.6	110000	8	AY506529	Continuation (2 of	c 499	22.8	45.6	179656	2	AC090473	Homo sapi
427	22.8	45.6	110445	2	AC141693	AC141693 Apis mell	c 500	22.8	45.6	180559	9	AL645922	Human DNA
428	22.8	45.6	111426	2	AC141693	AC141693 Apis mell							
429	22.8	45.6	111704	2	AC146847	AC146847 Ornithorh							
430	22.8	45.6	112930	9	AF326972	AF326972 Homo sapi							
431	22.8	45.6	113689	8	AC139707	AC139707 Medicago							
432	22.8	45.6	114089	10	AL807770	AL807770 Mouse DNA							
433	22.8	45.6	117995	10	AL929179	AL929179 Mouse DNA							
434	22.8	45.6	120480	9	AP000806	AP000806 Homo sapi							
435	22.8	45.6	122274	9	AC105359	AC105359 Homo sapi							
436	22.8	45.6	122792	2	CR388202	CR388202 Homo sapi							
437	22.8	45.6	122842	2	AC129474	AC129474 Homo sapi							
438	22.8	45.6	123837	2	AC013661	AC013661 Homo sapi							
439	22.8	45.6	124218	2	AC149131	AC149131 Medicago							
440	22.8	45.6	128305	2	AC132859	AC132859 Mus muscu							
441	22.8	45.6	129149	9	HS10D11	Z98044 Human DNA s							
442	22.8	45.6	130740	8	AC144343	AC144343 Medicago							
443	22.8	45.6	131266	10	AL669868	AL669868 Mouse DNA							
444	22.8	45.6	132323	9	AC011473	AC011473 Homo sapi							
445	22.8	45.6	137658	9	HS1022J11	AL049765 Human DNA							
446	22.8	45.6	137739	9	HS99419	AL034554 Human DNA							
447	22.8	45.6	141079	9	HSR218C14	AL121894 Human DNA							
448	22.8	45.6	141319	8	AP004457	AP004457 Oryza sat							
449	22.8	45.6	141675	2	AP000405	AP000405 Homo sapi							
450	22.8	45.6	142334	2	AC073185	AC073185 Homo sapi							
451	22.8	45.6	143212	9	AF106564	AF106564 Homo sapi							
452	22.8	45.6	143372	9	AL137847	AL137847 Human DNA							
453	22.8	45.6	147098	9	AC008651	AC008651 Homo sapi							
454	22.8	45.6	148316	2	AL356483	AL356483 Homo sapi							
455	22.8	45.6	148820	10	AC123926	AC123926 Mus muscu							
456	22.8	45.6	153464	9	AL844853	AL844853 Human DNA							
457	22.8	45.6	153860	10	AC101983	AC101983 Mus muscu							

ALIGNMENTS

BD078129 50 bp DNA linear PAT 27-AUG-2002
Modulator of DNA cytosine-5 methyltransferase and method of using the same.

BD078129
BD078129
BD078129.1 GI:22623732
JP 2001514862-A/13.
unidentified
unidentified
unclassified.
1 (bases 1 to 50)
Reich,N.O. and Flynn,J.
Modulator of DNA cytosine-5 methyltransferase and method of using the same

Patent: JP 2001514862-A 13 18-SEP-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Unidentified
PN JP 2001514862-A/13
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH, JAMES FLYNN
PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00
CC Strandedness: Double;


```

CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
FH Key Location/Qualifiers
FT source 1..50
FT /organism='Unidentified'.

FEATURES
    source
        Location/Qualifiers
            1..50
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'

ORIGIN
Query Match 100.0%; Score 50; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTACCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCCAC 50
Db |||||
1 CCTACCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCCAC 50

RESULT 2
BD078125/c
LOCUS
DEFINITION
Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
ACCESSION
BD078125
VERSION
BD078125.1 GI:22623728
KEYWORDS
JP 2001514862-A/9.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 30)
AUTHORS
Reich,N.O. and Flynn,J.
TITLE
Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL
Patent: JP 2001514862-A 9 18-SEP-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT
OS Unidentified
PN JP 2001514862-A/9
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
FH Key Location/Qualifiers
FT source 1..30
FT /organism='Unidentified'.

FEATURES
    source
        Location/Qualifiers
            1..30
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'

ORIGIN
Query Match 60.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 40
Db |||||
1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30

RESULT 4
BD005937
LOCUS
DEFINITION
Caulobacter crescentus CBI5 section 263 of 359 of the complete
genome.
ACCESSION
AE005937 AE005673
VERSION
AE005937.1 GI:13424291
KEYWORDS
JP 2001514862-A/10.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 10805)
AUTHORS
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Hart,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
PUBMED
11259647
REFERENCE
2 (bases 1 to 10805)
AUTHORS
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,

```

```

the same.
BD078126
VERSION
BD078126.1 GI:22623729
KEYWORDS
JP 2001514862-A/10.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 30)
AUTHORS
Reich,N.O. and Flynn,J.
TITLE
Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL
Patent: JP 2001514862-A 10 18-SEP-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT
OS Unidentified
PN JP 2001514862-A/10
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
FH Key Location/Qualifiers
FT source 1..30
FT /organism='Unidentified'.

FEATURES
    source
        Location/Qualifiers
            1..30
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'

ORIGIN
Query Match 60.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 40
Db |||||
1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30

RESULT 4
BD005937
LOCUS
DEFINITION
Caulobacter crescentus CBI5 section 263 of 359 of the complete
genome.
ACCESSION
AE005937 AE005673
VERSION
AE005937.1 GI:13424291
KEYWORDS
JP 2001514862-A/10.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 10805)
AUTHORS
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Hart,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
PUBMED
11259647
REFERENCE
2 (bases 1 to 10805)
AUTHORS
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,

```

Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathavan,J.,
Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
Fraser,C.M.

Direct Submission

Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES

Location/Qualifiers	
1..10805	
/organism="Caulobacter crescentus CB15"	
/mol_type="genomic DNA"	
/strain="CB15"	
/db_xref="taxon:190650"	
complement(161..391)	
/gene="CC2705"	
complement(161..391)	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="transcriptional regulator, Cro/CI family"	
/protein_id="AAK24670.1"	
/db_xref="GI:13424292"	
/translation="MRRNMRLRVVGANIRALREARGVPQDELAHVAIEIHTYLSGI ENKKNISLRVLRLAALVWSEALVRRKNQ"	
complement(469..693)	
/gene="CC2706"	
complement(469..693)	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="transcriptional regulator, Cro/CI family"	
/protein_id="AAK24671.1"	
/db_xref="GI:13424293"	
/translation="MRPFLMDPLAVVGANVTLREARGLPQDELALHAGVHATYLSGI ETGKNLTKVLERLASALRVPETDLVRR"	
819..1337	
/gene="CC2707"	
819..1337	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="RNA polymerase sigma-70 factor, ECF subfamily"	
/protein_id="AAK24672.1"	
/db_xref="GI:13424294"	
/translation="MSPLDGQITAPDASPSDGTPLADAVLRHLDMLTALLARYPQ DADDTAQOTWLRVHDYRPPRVPVRSRLLARIALNLAASRFKSSREIPTDPDHAPLA GRVIDASQOAVLPEQMLALPPKLURDVFVNLHVRLTVREIAQLRGISVKAVERMT QAIARCAQMIRG"	
1350..2423	
/gene="CC2708"	
1350..2423	
/note="CC2708"	
/note="similar to SP:P23485 GB:M63115 PID:145931 PID:537133 GB:U00096; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="transcriptional regulator, putative"	
/protein_id="AAK24673.1"	
/db_xref="GI:13424295"	
/translation="MIDQAKAANQPGGAPSGARDRRARDEAANVYKMSQGVSN ADLAFFAWRENSDAAYTRIEALTTSVRALADDPRIQAIATRAEAPRGVAKLR LQQAAGRTAGVGAALVAAMVIGVLLGVOQQYTRTQIGERRAVSLDDGSTIELNTN SOVRVRLSKTORSLLDRQAMFIVAHDAARPFVVTAGDTSVRAIGTRFEVYRAGAV RYTLAEGRVQYSKAGRDGGGGERSTFVILNAGQVDMAPAPSPVADVAAVT GWTRGLTFKQASLAEAVAENVYRERQVVLGPGTPAQRVNGEFDAGDTETTFARGMA VLLDLKTVARPDGGTLEAASPAV"	
2546..5632	
/gene="CC2709"	
2546..5632	

gene	gene
CDS	CDS


```
/note="match: cDNAs: Em:AB018315 Em:AB017026 Em:AK001079
Em:AB014604 Em:DI7006 Em:L29468
match: ESTs: Em:BE304813 Em:BE162519 Em:BE334880
Em:AV600093 Em:BE378914 Em:AW604852 Em:BE044221
Em:BE841635 Em:AV307639 Em:BE146176 Em:T60859 Em:AL121488
Em:AA349435 Em:AA628418 Em:BE165645 Em:BE883946
Em:AA339145 Em:A1548374 Em:BE671184 Em:BE464418
Em:AW784739 Em:AW935174 Em:A1138914 Em:BE044179 Em:T60886
Em:A1904635 Em:AW888820 Em:BE840640 Em:AW374601
Em:AW245234 Em:BE36988 Em:AI393982 Em:AW296964
Em:A1904905 Em:A1904903 Em:AA643501 Em:BE159704
Em:AA885187 Em:A1904901 Em:W23476 Em:AA157535 Em:AA301257
Em:AW168016 Em:AA871996 Em:H78948 Em:H47469 Em:T92296
Em:T8513740 Em:A1718509 Em:T84285 Em:AI339588 Em:A1279592
Em:A1339740 Em:AI005302 Em:AA993511 Em:A1219440
Em:AW809568 Em:A1250872 Em:AW970187 Em:BE169467
Em:AW166207 Em:A1904832 Em:AI935129 Em:AW589610
Em:A1806829 Em:AW468681 Em:AW242104 Em:A1864343
Em:A1904831 Em:A1492555 Em:A1904911 Em:N95070 Em:AW249708
Em:AW418677 Em:A1859944 Em:A1346158 Em:AW614781
Em:A1355152 Em:BE735608 Em:A1904910 Em:AW959411 Em:N29972
Em:AW8768 Em:A1800827 Em:N51060 Em:AW175867 Em:T92242
Em:AW439520 Em:AA862307 Em:A1910432 Em:AA448644
Em:BE816019 Em:A1378973 Em:AA757012 Em:A1831062
Em:AA604829 Em:BE818182 Em:A1686611 Em:A1566514
Em:A1904852 Em:AA448740 Em:AA437158 Em:N55701 Em:A1452737
Em:A1935875 Em:A1233219 Em:AW602161 Em:A5599578 Em:N34529
Em:AW167999 Em:M7888 Em:A1142731 Em:AW340478 Em:AA860591
Em:A1278471 Em:AW265178 Em:AW511636 Em:T90208 Em:T92252
Em:A1146325 Em:BE808784 Em:A149667 Em:A1305178
Em:A1753638 Em:AA309462 Em:A1689455 Em:AW081922"
/evidence=not_experimental
2999. 3208
/note="7 copies 30 mer 61% conserved"
3012. 3179
/note="7 copies 24 mer 68% conserved"
3013. 3180
/note="28 copies 6 mer cgggg 61% conserved"
3136. 3165
/gene="BA157P1.3"
/note="Single clone region. Assembly consistent with
restriction digest."
4277. 4577
/note="AluSq repeat: matches 1. .302 of consensus"
4649. 4743
/note="12 repeat: matches 2645. .2750 of consensus"
4825. 5058
/note="MIR repeat: matches 2. .252 of consensus"
5601. 5788
/note="AluY repeat: matches 129. .294 of consensus"
5874. 6203
/gene="BA157P1.3"
/note="match: GSS: Em:AAQ069568"
7847. 7916
/note="5 copies 14 mer 88% conserved"
7982. 8095
/note="MER5A repeat: matches 52. .165 of consensus"
Complement(8837. 9343)
/note="match: GSS: Em:AAQ277251"
Complement(8889. 9350)
/note="match: GSS: Em:AAQ311219"
Complement(9023. 9388)
/note="match: GSS: Em:AAQ392442"
9523. 9817
/note="AluX repeat: matches 1. .300 of consensus"
10658. 10824
/note="MER3 repeat: matches 14. .203 of consensus"
12083. 12387
/note="AluSg repeat: matches 1. .307 of consensus"
12568. 12878
/note="AluX repeat: matches 3. .312 of consensus"
13102. 13422
/note="AluJo repeat: matches 1. .312 of consensus"
```

```
repeat_region 13986. 14298
/note="AluX repeat: matches 1. .311 of consensus"
misc_feature complement(14510. 14953)
/note="match: GSS: Em:AAQ891002"
repeat_region 14546. 14615
/note="12 repeat: matches 2634. .2710 of consensus"
misc_feature 14778. 15212
Query Match 53.2%; Score 26.6; DB 9; Length 141056;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

QY 1 CCTACCCACCTTGATCTTGCCTGCGCCCTTGATTCCTCCACCTCCA 49

Db 133445 CCACCCCACTCTCCCTGCGCCCACTCGCAATCCACCTCCA 133397

RESULT 6

```
AC129725 152806 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-422P4, WORKING DRAFT SEQUENCE.
DEFINITION AC129725.3 GI:25074173
ACCESSION HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION Rattus norvegicus (Norway rat)
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
```

REFERENCE

```
1 (bases 1 to 152806)
Munzy,D.Marie., Metzker,M.Iee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
Anyalbechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.I., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,J.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoelameh,O., Okwuon,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pul-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sander,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabok,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
```

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlarczyk, R., Wooden, R., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission
 2 (bases 1 to 152806)
 Worley, K.C.

Direct Submission
 Submitted (02-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 152806)
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23665108. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KHP
 Center clone name: CH230-422F4
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 139811 bases at least Q40
 Consensus quality: 140863 bases at least Q30
 Consensus quality: 141480 bases at least Q20
 Estimated insert size: 142316; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and the accession number will be preserved.
 * 1 152806: contig of 152806 bp in length.
 Location/Qualifiers
 1..152806
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-422F4"
 1..1221
 /note="wgs end_extension
 clone_end:77"
 6494..7146
 /note="clone_boundary
 clone_end:77"

misc_feature
 misc_feature

FEATURES
 source

site:
 end_sequence:BZ119845"
 143925..144873
 /note="clone_boundary
 clone_end:Sp6"
 site:
 end_sequence:BZ119846"
 145455..146987
 /note="wgs_end_extension
 clone_end:Sp6"
 151655..152806
 /note="wgs_end_extension
 clone_end:Sp6"

misc_feature
 misc_feature
 misc_feature

ORIGIN

Query Match 53.2%; Score 26.6; DB 2; Length 152806;
 Best Local Similarity 71.4%; Pred. No. 25;
 Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCTTGCCCGCCCTTGAAATTCACCACTCCAC 50
 Db 130 CCACCTACCTCCATCCATCACCTCCCATTCATTTACCATCTCCAC 178

RESULT 7
 AC136047/c
 AC136047/1
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-190M11, *** SEQUENCING IN PROGRESS
 *** 62 unordered pieces.
 AC136047
 Rattus norvegicus (Norway rat)
 HTG: HTGS PHASE1.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 194785)
 Muzny, D.Marie., Metzker, M.Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., Mcleod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpuneagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,

Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C. D., Smajd, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,
 Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished
 2 (bases 1 to 194785)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (29-Oct-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KDDW

Center clone name: CH230-190M11

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 146031 bases at least Q40

Consensus quality: 150884 bases at least Q30

Consensus quality: 154819 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1019: contig of 1019 bp in length

* 1020 1119: gap of unknown length

* 1120 2276: contig of 1157 bp in length

* 2277 2376: gap of unknown length

* 2377 3542: contig of 1166 bp in length

* 3543 3642: gap of unknown length

* 3643 5191: contig of 1549 bp in length

* 5192 5291: gap of unknown length

* 5292 6410: contig of 1119 bp in length

* 6411 6510: gap of unknown length

* 6511 7615: contig of 1105 bp in length

* 7616 7715: gap of unknown length

* 7716 9116: contig of 1401 bp in length

* 9117 9216: gap of unknown length

* 9217 10427: contig of 1211 bp in length

* 10428 10527: gap of unknown length

* 10528 11989: contig of 1462 bp in length

* 11990 12089: gap of unknown length

* 12090 13249: contig of 1160 bp in length

* 13250 13349: gap of unknown length

* 13350 14820: contig of 1471 bp in length

* 14821 14920: gap of unknown length

* 14921 16065: contig of 1145 bp in length

* 16066 16165: gap of unknown length

* 16166 17869: contig of 1704 bp in length

* 17870 17969: gap of unknown length

17970 19080: contig of 1111 bp in length

* 19081 19180: gap of unknown length

* 19181 20915: contig of 1735 bp in length

* 20916 21015: gap of unknown length

* 21016 22475: contig of 1460 bp in length

* 22476 22575: gap of unknown length

* 22576 24020: contig of 1445 bp in length

* 24021 24120: gap of unknown length

* 24121 25142: contig of 1022 bp in length

* 25143 26666: gap of unknown length

* 26667 26766: contig of 1424 bp in length

* 26767 28874: gap of unknown length

* 28875 28974: gap of unknown length

* 28975 31189: contig of 2215 bp in length

* 31190 31289: gap of unknown length

* 31290 33543: contig of 2254 bp in length

* 33544 33643: gap of unknown length

* 33644 34692: contig of 1049 bp in length

* 34693 34792: gap of unknown length

* 34793 36704: contig of 1912 bp in length

* 36705 36804: gap of unknown length

* 36805 38966: contig of 2162 bp in length

* 38967 39066: gap of unknown length

* 39067 40705: contig of 1639 bp in length

* 40706 40805: gap of unknown length

* 40806 42258: contig of 1453 bp in length

* 42259 42358: gap of unknown length

* 42359 43854: contig of 1496 bp in length

* 43855 43954: gap of unknown length

* 43955 45800: contig of 1846 bp in length

* 45801 45900: gap of unknown length

* 45901 48110: contig of 2210 bp in length

* 48111 48210: gap of unknown length

* 48211 50389: contig of 2179 bp in length

* 50390 50489: gap of unknown length

* 50490 53560: contig of 3071 bp in length

* 53561 53660: gap of unknown length

* 53661 55661: contig of 2048 bp in length

* 55662 55808: gap of unknown length

* 55809 58018: contig of 3210 bp in length

* 58019 59118: gap of unknown length

* 59119 61871: contig of 2753 bp in length

* 61872 61971: gap of unknown length

* 61972 65183: contig of 3212 bp in length

* 65184 65283: gap of unknown length

* 65284 68466: contig of 3183 bp in length

* 68467 68566: gap of unknown length

* 68567 71348: contig of 2782 bp in length

* 71349 71448: gap of unknown length

* 71449 74855: contig of 3407 bp in length

* 74856 74955: gap of unknown length

* 74956 78006: contig of 3051 bp in length

* 78007 78106: gap of unknown length

* 78107 80552: contig of 2446 bp in length

* 80553 80652: gap of unknown length

* 80653 84966: contig of 4314 bp in length

* 84967 85066: gap of unknown length

* 85067 87929: contig of 2863 bp in length

* 87930 88029: gap of unknown length

* 88030 91972: contig of 3943 bp in length

* 91973 92072: gap of unknown length

* 92073 96600: contig of 4528 bp in length

* 96601 96700: gap of unknown length

* 96701 99667: contig of 2367 bp in length

* 99668 99767: gap of unknown length

* 99768 103569: contig of 3802 bp in length

* 103570 103669: gap of unknown length

* 103670 107511: contig of 3842 bp in length

* 107512 107612: gap of unknown length

* 107613 114112: contig of 6501 bp in length

* 114113 114212: gap of unknown length

* 114213 120068: contig of 5856 bp in length

```

* 120069 120168: gap of unknown length
* 120169 123964: contig of 3796 bp in length
* 123965 124064: gap of unknown length
* 124065 129323: contig of 5259 bp in length
* 129324 129423: gap of unknown length
* 129424 135042: contig of 5619 bp in length
* 135043 135142: gap of unknown length

Query Match      53.2%; Score 26.6; DB 2; Length 194785;
Best Local Similarity 71.4%; Pred. NO. 25;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGCATCTTCCCGCCCTTGATTCACCACTCCAC 50
DB 40979 CCACCTACCTCATCCATCCATCCCTCCATTCATTTACCCATCTCCAC 40931

RESULT 8
AC105637
LOCUS      239910 bp      DNA      linear      HTG 08-OCT-2002
DEFINITION Rattus norvegicus clone CH230-140K9, *** SEQUENCING IN PROGRESS
ACCESSION AC105637
VERSION    3
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus
REFERENCE  1 (bases 1 to 239910)
           Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
           Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
           Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
           Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
           Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
           Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
           Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
           Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
           Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
           Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
           Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K.,
           Draper, H., Dugan-Kocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
           Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
           Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
           Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
           Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,
           Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hernandez, K.,
           Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
           Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
           Hollins, B., Howells, S., Hui, S., Hume, J., Idlebird, D., Jackson, A.,
           Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
           Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Liu, J.,
           Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
           Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
           Lorenshewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J.,
           Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A.,
           Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
           Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
           Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
           Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
           Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
           Nwankweli, O., Okwuonu, G., Olarnpunaagoon, A., Pal, S., Parks, K.,
           Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
           Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,
           Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
           Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
           Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
           Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
           Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
           Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sooa, J.,
           Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
           Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

```

```

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 239910)
Worley, K. C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239910)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNKM
Center clone name: CH230-140K9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 190013 bases at least Q40
Consensus quality: 194473 bases at least Q30
Consensus quality: 197802 bases at least Q20
Estimated insert size: 211687; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 238338: contig of 238338 bp in length
* 238339 238438: gap of unknown length
* 238439 239910: contig of 1472 bp in length.
Location/Qualifiers
1. 239910
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-140K9"
18349..20312
/misc_feature
/note="wgs contig"
100522..102076
/misc_feature
/note="wgs_contig"

```

FEATURES
source

misc_feature
misc_feature

```

misc_feature 179779..182739
              /note="wgs_contig"

ORIGIN

Query Match      53.2%; Score 26.6; DB 2; Length 239910;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATTCCTGCCCGCCCTTGAATTCACCAACCTCCAC 50
Db 136233 CCACCTACCCCTCATCATCACCTCCCATTCATTTACCCATCCTCCAC 136281

RESULT 9
AC094206 AC094206 241243 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-3A15, WORKING DRAFT SEQUENCE, 4
DEFINITION AC094206
ACCESSION AC094206.6 GI:30467689
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 241243)
Muzny, D. Marie., Metzker, M. Lee., Abravzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Guebragorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, W.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louiseged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Platter, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Vatas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

```

```

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 241243)
Worley, K. C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241243)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24819491.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAET
Center clone name: CH230-3A15
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 226414 bases at least Q40
Consensus quality: 229449 bases at least Q30
Consensus quality: 231607 bases at least Q20
Estimated insert size: 233457; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 54907: contig of 54907 bp in length
* 54908 55007: gap of unknown length
* 55008 237646: contig of 182639 bp in length
* 237647 237746: gap of unknown length
* 237747 239762: contig of 2016 bp in length
* 239763 239862: gap of unknown length
* 239863 241243: contig of 1381 bp in length.
          Location/Qualifiers
            1..241243
              /organism="Rattus norvegicus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10116"
              /clones="CH230-3A15"
            1..1261
              /note="wgs_contig"
            53814..54907
              /note="wgs_contig"

```

```

FEATURES
source
misc_feature
misc_feature

```


variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-396N6 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACes3.6.

Location/Qualifiers

1. 217726

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="2"

/clone="RP23-396N6"

/clone_lib="RPCI-23"

33; Similarity 52.8%; Score 26.4; DB 10; Length 217726;

33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

7 CACCTGATCTTGGCCCGCCCTTGAAATCCCAACCTCCAC 50

|||||

7 CACCTGCTTCTTGGCCCGCCCTTCCCTCCCTCCCTCCCTCC 28120

|||||

AP005019 98501 bp DNA linear PRI 30-MAR-2002

Homo sapiens genomic DNA, chromosome 11q clone:RP11-169D4, complete sequences.

AP005019

AP005019.1 GI:19879835

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Homo sapiens genomic DNA

Published Only in Database (2002)

2 (bases 1 to 98501)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (28-MAR-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)

Location/Qualifiers

1. 98501

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="RP11-169D4"

```

Query Match          52.0%; Score 26; DB 9; Length 98501;
Best Local Similarity 70.0%; Pred.No. 43;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTCGATCTCTGCGCGCCCTTGAATTCACACCTCCAC 50
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22133 CCTTCCCTCCCTCTCTCCCTCCACCTCTCTCCCTTCCACCTCC 22084

RESULT 13
AC128662
LOCUS Mus musculus BAC clone RP24-440E3 from chromosome 6, complete
DEFINITION Mus musculus BAC clone RP24-440E3 from chromosome 6, complete
sequence.
ACCESSION AC128662
VERSION AC128662.4 GI:28867173
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 134729)
TITLE The sequence of Mus musculus BAC clone RP24-440E3
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 134729)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 134729)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 134729)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 134729)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 134729)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 6, 2003 this sequence version replaced gi:28604183.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0440E03
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES	source
	Location/Qualifiers
	1..134729
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="6"
	/map="6"
	/clone="RP24-440E3"
	/clone_lib="RPCI-24"
	60..135
repeat_region	/rpt_family="ID"
repeat_region	913..1121
repeat_region	/rpt_family="B4"
repeat_region	1317..1560
repeat_region	/rpt_family="MER2_type"
repeat_region	5475..5617
repeat_region	/rpt_family="Alu"
repeat_region	5618..5764
repeat_region	/rpt_family="Alu"
repeat_region	5812..6007
repeat_region	/rpt_family="L1"
repeat_region	6052..6170
repeat_region	/rpt_family="Alu"
repeat_region	6188..6723
repeat_region	/rpt_family="L1"
repeat_region	6722..7168
repeat_region	/rpt_family="L1"
repeat_region	7200..7358
repeat_region	/rpt_family="B4"
repeat_region	10550..10703
repeat_region	/rpt_family="B4"
repeat_region	11010..11050
repeat_region	/rpt_family="B2"
repeat_region	11051..11195
repeat_region	/rpt_family="Alu"
repeat_region	11196..11364
repeat_region	/rpt_family="B2"
repeat_region	11397..11445
repeat_region	/rpt_family="B2"
repeat_region	11529..11583
repeat_region	/rpt_family="MaLR"
repeat_region	14286..14601
repeat_region	/rpt_family="MaLR"
repeat_region	14884..14951
repeat_region	/rpt_family="L1"
repeat_region	15815..15924
repeat_region	/rpt_family="Alu"
repeat_region	16152..16283
repeat_region	/rpt_family="Alu"
repeat_region	16376..16604
repeat_region	/rpt_family="B2"
repeat_region	16658..16807
repeat_region	/rpt_family="B4"
repeat_region	17029..17111
repeat_region	/rpt_family="4.5SRNA"
repeat_region	18197..18323
repeat_region	/rpt_family="B4"
repeat_region	18469..18612
repeat_region	/rpt_family="Alu"

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Keils, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Manning, J., Matthews, C., McCarthy, M., MacDonald, P., Mayor, J., Minning, J., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataranan, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 9, 2003 this sequence version replaced gi:30985094.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li4251
Center clone name: 54_N_13

FEATURES

source

Location/Qualifiers

1..251756
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-54N13"
/clone_lib="RPCI-23 Female Mouse BAC"

misc_feature

1..6
/note="clone boundary
clone end:SP6
site:EcoRI"

repeat_region

complement(9..288)
/rpt_family="RMER12"

repeat_region

2620..2648
/rpt_family="(ATG)n"

repeat_region

complement(2730..3077)
/rpt_family="ORR1D"

repeat_region

3229..3349
/rpt_family="ORR1D"

repeat_region

4652..5220
/rpt_family="L1"

repeat_region

5224..5324
/rpt_family="Lx2"

repeat_region

complement(5606..5713)
/rpt_family="L1_MM"

repeat_region

5712..6084
/rpt_family="L1_MM"

repeat_region

7442..7469
/rpt_family="AT_rich"

repeat_region

7766..7798
/rpt_family="(CATATA)n"

repeat_region

7856..7917
/rpt_family="(TG)n"

repeat_region

complement(7918..8031)
/rpt_family="RSINE1"

repeat_region

complement(8692..8963)
/rpt_family="RMER15"

repeat_region

complement(9130..9989)
/rpt_family="L1"

repeat_region

10160..10428
/rpt_family="MERVL_LTR"

repeat_region 10344..10624
/rpt_family="MT2B"
repeat_region complement(11226..11321)
/rpt_family="Lx3"
repeat_region complement(11327..12562)
/rpt_family="Lx3"
repeat_region 12563..13468
/rpt_family="RMER17A2"
repeat_region complement(13469..16671)
/rpt_family="Lx3"
repeat_region complement(16663..17151)
/rpt_family="L1V12"
repeat_region complement(17303..17436)
/rpt_family="L1F"
repeat_region 17931..17960
/rpt_family="(CAA)n"
repeat_region complement(18620..18983)
/rpt_family="RLTRET_N_MM"
repeat_region 20103..20132
/rpt_family="(TTG)n"
repeat_region 20186..20233
/rpt_family="RMER12"
repeat_region complement(20344..20565)
/rpt_family="URR1B"
repeat_region complement(20569..20990)
/rpt_family="Lx8"
repeat_region 21616..21652
/rpt_family="(TTTTA)n"
repeat_region 21837..21974
/rpt_family="L1_MM"
repeat_region 28585..28657
/rpt_family="(TCTCTG)n"
repeat_region 28655..28686
/rpt_family="(TG)n"
repeat_region complement(28718..28793)
/rpt_family="tRNA-Ala-GCA"
repeat_region complement(29278..29585)
/rpt_family="Lx9"
repeat_region complement(29596..29767)
/rpt_family="L1M2"
repeat_region 29861..29904
/rpt_family="T-rich"
repeat_region 30003..30180
/rpt_family="(TTTC)n"
repeat_region complement(31570..31846)
/rpt_family="MLT1"
repeat_region 32848..32811
/rpt_family="RSINE1"
repeat_region 32821..32991
/rpt_family="(CAGAGA)n"
repeat_region complement(33036..33175)
/rpt_family="Lx"
repeat_region 33173..33518
/rpt_family="L1_MM"
repeat_region 33679..33754
/rpt_family="MIR"
repeat_region 34978..35010
/rpt_family="AT_rich"
repeat_region 35109..35136

Query Match 52.0%; Score 26; DB 10; Length 251756;

Best Local Similarity 76.2%; Pred. No. 40;

Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 CCCTGGATCCTTGGCCCCCGCCCTTGAATCCCAACCCCTCCAC 50

Db 127191 CCCTGTTCCCTGGCCCCCGCCCTTGAATTCATCCCTCCTC 127150

RESULT 15

AB115779

LOCUS

DEFINITION Bos taurus CLDN2 mRNA for claudin 2, complete cds, tissue_type:

1032 bp mRNA linear

MAM 18-NOV-2003

```

kidney.
ACCESSION AB115779
VERSION AB115779.1 GI:38175210
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1
REFERENCE Ohta,H., Takiguchi,M. and Inaba,M.
AUTHORS Localization of claudin proteins in bovine kidneys
TITLE Unpublished
JOURNAL
2 (bases 1 to 1032)
AUTHORS Ohta,H., Takiguchi,M. and Inaba,M.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2003) Hiroshi Ohta, Hokkaido University Graduate
School of Veterinary Medicine, Laboratory of Molecular Medicine;
N18 W9, Sapporo, Hokkaido 060-0818, Japan
(E-mail:h-ohata@vetmed.hokudai.ac.jp, Tel:81-11-706-5276(ex.5276),
Fax:81-11-706-5240)
FEATURES
source Location/Qualifiers
1..1032
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="kidney"
1..1032
/gene="CLDN2"
22..714
/gene="CLDN2"
/note="right junction-associated protein"
/codon_start=1
/evidence=experimental
/product="claudin 2"
/db_xref="GI:38175211"
/db_xref="GI:38175211"
/translation="WASLGQLGVYVGLGLLGTVIAMLPSTWTSYVVGASIVTAV
GFSKLMWECATHSTGTCQDITSTMLGPDIAQAQAMVTSSSLACTIVSVMG
RTEVFPQSRADKAVVGVGVFLLGLGFTPVANLHGLRDFYSLVPDPMKFEI
GKALVLIHSLSLFIAGIFLCFSCSPQGNRSNYDAYOAOQLATRSSPRQAPKKG
SEFNYSYSLTGYV"
ORIGIN
Query Match 51.2%; Score 25.6; DB 4; Length 1032;
Best Local Similarity 69.4%; Pred. No. 91;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 2 CTACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCCACCTCCAC 50
Db ||||| ||||| ||||| ||||| ||||| ||||| |||||
950 CTTGGCACCCTCCACCCACCCAGCCCTCAATATCCCAACCCCTCAAC 998
RESULT 16
AC117870 204144 bp DNA linear HTG 19-SEP-2002
LOCUS Rattus norvegicus clone CH230-253N3, *** SEQUENCING IN PROGRESS
DEFINITION ***. 2 unordered pieces.
AC117870 AC117870.6 GI:23194665
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 204144)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

```

```

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falla,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,T., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idelbird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenschwah,L., Loulsegged,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazb,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,I., Rojas,A., Rose,W., Rose,R., Ruiz,S., Shen,H.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Tabor,P., Taylor,K.,
Valas,R., Vera,V., Villalana,B., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 204144)
Worley,K.C.
Direct Submission
Submitted (11-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 204144)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21746235.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GRZK
Center Clone name: CH230-253N3

```

```
----- Summary Statistics
Consensus program: Phrap; version 0.990329
Consensus quality: 186062 bases at least Q40
Consensus quality: 187534 bases at least Q30
Consensus quality: 188573 bases at least Q20
Estimated insert size: 202328; sum-of-coverage: 4x in Q20 bases; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 131474: contig of 131474 bp in length
* 131475 131574: gap of unknown length
* 131575 204144: contig of 72570 bp in length.
* Location/Qualifiers
  . 204144
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-253N3"
  misc_feature 131575..132927
    /note="wgs_contig"
  misc_feature 138146..139255
    /note="wgs_contig"
ORIGIN
Query Match 51.2%; Score 25.6; DB 2; Length 204144;
Best Local Similarity 70.8%; Pred. No. 57;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 CCTACCCACCCGCGATCCTTGCCCGCCCTTGATTCACCCCTCC 48
Db 155707 CCTCCCTCCCGCCCTCTTCTCCTCCCTTCTCCCTCCCTCC 155754

RESULT 17
AC129955/c
LOCUS
DEFINITION Bos taurus clone RP42-225J6, *** SEQUENCING IN PROGRESS ***, 4
unordered pieces.
ACCESSION AC129955
VERSION AC129955.8 GI:45736837
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 204513)
Barnia.N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camaraeta,J., Chang,J., Chararo,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo.A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Submitted (25-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
ReMar 25, 2004 This sequence version replaced gi:41393342.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@broad.mit.edu
----- Project Information
Center project name: L24576
Center clone name: 225_J_6
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 73851: contig of 73851 bp in length
* 73852 73951: gap of 100 bp
* 73952 195199: contig of 121248 bp in length
* 195200 195299: gap of 100 bp
* 195300 203734: contig of 8435 bp in length
* 203735 203834: gap of 100 bp
* 203835 204513: contig of 679 bp in length.
* Location/Qualifiers
  1..204513
  /organism="Bos taurus"
  /mol_type="genomic DNA"
  /db_xref="taxon:9913"
  /clone="RP42-225J6"
  /clone_lib="RPC1-42 Bovine BAC"
FEATURES
source
ORIGIN
Query Match 51.2%; Score 25.6; DB 2; Length 204513;
Best Local Similarity 70.8%; Pred. No. 57;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```


LOCUS	AC129959	211988 bp	DNA	linear	MAM 24-JAN-2003			
DEFINITION	Bos taurus, clone RP42-518P7, complete sequence.							
ACCESSION	AC129959							
VERSION	AC129959.6	GI:27884899						
KEYWORDS	HTG.							
SOURCE	Bos taurus (cow)							
ORGANISM	Bos taurus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.							
AUTHORS	Birren,B., Nussbaum,C. and Lander,E.							
TITLE	1 (bases 1 to 211988)							
JOURNAL	Bos taurus, clone RP42-518P7							
REFERENCE	Unpublished							
AUTHORS	2 (bases 1 to 211988)							
REFERENCE	Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chararo,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.							
	Zembek,L., Zimmer,A. and Zody,M.							
Direct Submission								
JOURNAL	Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA							
REFERENCE	3 (bases 1 to 211988)							
AUTHORS	Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chararo,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.							
Direct Submission								
JOURNAL	Submitted (03-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA							
REFERENCE	4 (bases 1 to 211988)							
AUTHORS	Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hagos,B., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,							

Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	Direct Submission	
Submitted (24-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Research	320 Charles Street, Cambridge, MA 02141, USA
On Jan 24, 2003 this sequence version replaced gi:27476173.	COMMENT	
All repeats were identified using RepeatMasker:		
Smit, A.F.A. & Green, P. (1996-1997)		
http://ftp.genome.washington.edu/RM/RepeatMasker.html		
----- Genome Center		
Center: Whitehead Institute/ MIT Center for Genome Research		
Center code: WtBR		
Web site: http://www-seq.wi.mit.edu		
Contact: sequence_submissions@genome.wi.mit.edu		
----- Project Information		
Center Project name: L24459		
Center clone name: 518_P_7		

Location/Qualifiers		
1..211988		
/organism="Bos taurus"		
/mol_type="genomic DNA"		
/db_xref="taxon:9913"		
/clone="RP42-518P7"		
/clone_lib="RPC1-42 Bovine BAC"		
1..54		
/rpt_family="Bov-tA2"		
2..74		
/rpt_family="BOV-A2"		
309..570		
/rpt_family="BOV-A2"		
complement(583..786)		
/rpt_family="Bov-tA3"		
791..831		
/rpt_family="(TA)n"		
complement(838..917)		
/rpt_family="L1_Art"		
936..1030		
/rpt_family="D-repeat"		
942..1016		
/rpt_family="tRNA-Glu-GAG"		
complement(1049..1593)		
/rpt_family="L1MB2"		
complement(2968..3149)		
/rpt_family="MER20"		
3683..3710		
/rpt_family="AT rich"		
complement(3722..4647)		
/rpt_family="L1_BT"		
4987..5051		
/rpt_family="(GGGAA)n"		
complement(5216..5425)		
/rpt_family="L1_Art"		
5451..5639		
/rpt_family="Bov-B"		
5569..5648		
/rpt_family="BOV-A2"		
5779..6099		
/rpt_family="D-repeat"		
complement(7961..7992)		
/rpt_family="GC-rich"		
complement(9155..9367)		
/rpt_family="Bov-tA1"		
complement(9879..10093)		
/rpt_family="Bov-tA1"		
10278..10390		
/rpt_family="MON1"		
10441..10710		
/rpt_family="BOV-A2"		

Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	TITLE	JOURNAL	COMMENT
Submitted (24-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
On Jan 24, 2003 this sequence version replaced gi:27476173.	TITLE	JOURNAL	COMMENT
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html			
----- Genome Center	TITLE	JOURNAL	COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research			
Center code: WIER	TITLE	JOURNAL	COMMENT
Web site: http://www-seq.wi.mit.edu			
Contact: sequence_submissions@genome.wi.mit.edu	TITLE	JOURNAL	COMMENT
----- Project Information			
Center project name: L24459	TITLE	JOURNAL	COMMENT
Center clone name: 518_P_7			
----- Location/Qualifiers	TITLE	JOURNAL	COMMENT
1..211988			
/organism="Bos taurus"	TITLE	JOURNAL	COMMENT
/mol_type="genomic DNA"			
/db_xref="taxon:9913"	TITLE	JOURNAL	COMMENT
/clone="RP42-518P7"			
/clone_lib="RPCI-42 Bovine BAC"	TITLE	JOURNAL	COMMENT
1..54			
/rpt_family="Bov-tA2"	TITLE	JOURNAL	COMMENT
2..74			
/rpt_family="BOV-A2"	TITLE	JOURNAL	COMMENT
309..570			
/rpt_family="BOV-A2"	TITLE	JOURNAL	COMMENT
complement(583..786)			
/rpt_family="Bov-tA3"	TITLE	JOURNAL	COMMENT
791..831			
/rpt_family="(TA)n"	TITLE	JOURNAL	COMMENT
complement(838..917)			
/rpt_family="L1_Art"	TITLE	JOURNAL	COMMENT
936..1030			
/rpt_family="D-repeat"	TITLE	JOURNAL	COMMENT
942..1016			
/rpt_family="tRNA-Glu-GAG"	TITLE	JOURNAL	COMMENT
complement(1049..1593)			
/rpt_family="L1MB2"	TITLE	JOURNAL	COMMENT
complement(2968..3149)			
/rpt_family="MER20"	TITLE	JOURNAL	COMMENT
3683..3710			
/rpt_family="AT rich"	TITLE	JOURNAL	COMMENT
complement(3722..4647)			
/rpt_family="L1_BT"	TITLE	JOURNAL	COMMENT
4987..5051			
/rpt_family="(GGGAA)n"	TITLE	JOURNAL	COMMENT
complement(5216..5425)			
/rpt_family="L1_Art"	TITLE	JOURNAL	COMMENT
5451..5639			
/rpt_family="Bov-B"	TITLE	JOURNAL	COMMENT
5569..5648			
/rpt_family="BOV-A2"	TITLE	JOURNAL	COMMENT
5779..6099			
/rpt_family="D-repeat"	TITLE	JOURNAL	COMMENT
complement(7961..7992)			
/rpt_family="GC rich"	TITLE	JOURNAL	COMMENT
complement(9155..9367)			
/rpt_family="Bov-tA1"	TITLE	JOURNAL	COMMENT
complement(9879..10093)			
/rpt_family="Bov-tA1"	TITLE	JOURNAL	COMMENT
10278..10390			
/rpt_family="MON1"	TITLE	JOURNAL	COMMENT
10441..10710			
/rpt_family="BOV-A2"	TITLE	JOURNAL	COMMENT


```

repeat_region 11881..12323
/rpt_family="Bov-B"
repeat_region 13876..13989
/rpt_family="Bov-D"
repeat_region 14027..14110
/rpt_family="MON1"
repeat_region complement(14388..14577)
/rpt_family="Bov-ta2"
repeat_region 15109..15550
/rpt_family="LIMA9"
repeat_region 15125..15771
/rpt_family="L1_Art"
repeat_region 15785..15855
/rpt_family="L1_Art"
repeat_region complement(16706..16930)
/rpt_family="MLT1A"
repeat_region complement(16706..16788)
/rpt_family="MLT1C"
unsure complement(18970..18977)
/note="x30 qual SGL region"
repeat_region complement(20665..20769)
/rpt_family="LIMC5"
repeat_region 20885..20952
/rpt_family="MAR1"
unsure complement(21716..21729)
/note="single clone coverage"
repeat_region 23010..23139
/rpt_family="MON1"
repeat_region complement(23185..23448)
/rpt_family="Bov-A2"
repeat_region 23456..25122
/rpt_family="LIM2_orf2"
repeat_region 25123..25190
/rpt_family="Bov-D"
repeat_region 25259..26197
/rpt_family="L1A_Oc"
repeat_region 25501..26359
/rpt_family="LIMA9"
repeat_region complement(26359..26438)
/rpt_family="Bov-A2"
repeat_region complement(26368..26811)
/rpt_family="Bov-B"
repeat_region 26817..26991
/rpt_family="Bov-B"
repeat_region complement(26992..27061)
/rpt_family="Bov-ta1"
repeat_region 27063..27333
/rpt_family="Bov-A2"
repeat_region complement(27333..27491)
/rpt_family="Bov-ta1"
repeat_region 27491..27897
/rpt_family="Bov-B"
repeat_region complement(27935..28076)
/rpt_family="Bov-B"

Query Match 51.2%; Score 25.6; DB 4; Length 211988;
Best Local Similarity 70.8%; Pred. No. 57;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCTTGGCCCGCCCTTGAATTCGACCTCCCA 49
DB 115035 CCACTCTCCCTGGTCTTGGCCCGCCCTCTTCTCTCACTCTCCCA 115082

RESULT 20
AL645800/c AL645800 98204 bp DNA linear VRT 04-OCT-2003
LOCUS Zebrafish DNA sequence from clone XX-184L24 Contains three novel
DEFINITION genes and the sp1 gene for transcription factor sp1, complete
sequence.
ACCESSION AL645800
VERSION AL645800.5 GI:19847883
KEYWORDS HTG; sp1; transcription factor.

```

SOURCE ORGANISM

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 98204)

Beasley,H.

Direct Submission

Submitted (03-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Mar 29, 2002 this sequence version replaced gi:19031669.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep-XX-184L24 is

from a Incyte Zebrafish AB BAC library VECTOR: pBeloBAC11
This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the rare
assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.
This clone was isolated from the library FWT (EKK) Fish (Genome

Research, St. Louis, USA (now Incyte Genomics)) and provided by
Karl Heu (A. Thomas Look lab).

This sequence is the entire insert of clone XX-184L24 Clone-derived
zebrafish pUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be

submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson

lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC

Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

FEATURES source

Location/Qualifiers
1..98204

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="XX-184L24"

/clone_lib="Incyte-BAC"

551..738

/note="Dr000297 repeat: matches 7. .196 of consensus"

738..1073

/note="TDR2 repeat: matches 1. .352 of consensus"

complement(1074..1138)

/note="Dr000397 repeat: matches 1. .65 of consensus"

complement(1143..1328)

/note="Dr000387 repeat: matches 389. .581 of consensus"

1384..1394

/note="5.5 copies 2 mer TG 22% conserved"

complement(1609..1791)

/note="HE1 DRI repeat: matches 175. .372 of consensus"

2005..2025

/note="2.1 copies 10 mer ACTGCAAGAC 42% conserved"

2019..2029

repeat_region

```
repeat_region /note="2.2 copies 5 mer CAAGA 22% conserved"
2082. .2142
/note="Dr000351 repeat: matches 500. .558 of consensus"
repeat_region complement(2150. .2312)
/note="dir880 repeat: matches 3. .572 of consensus"
2332. .2343
/note="2.0 copies 6 mer ACATGG 24% conserved"
2349. .2358
/note="10.0 copies 1 mer T 20% conserved"
2748. .2760
/note="2.6 copies 5 mer TCTCC 26% conserved"
3475. .3484
/note="2.5 copies 4 mer TTAT 20% conserved"
3562. .3576
/note="7.5 copies 2 mer AT 21% conserved"
3774. .3789
/note="2.0 copies 8 mer ATAAATCC 32% conserved"
3803. .3950
/note="Dr000109 repeat: matches 6. .155 of consensus"
complement(3959. .4026)
/note="Dr000109 repeat: matches 660. .730 of consensus"
4008. .4074
/note="DNA9TAL_DR repeat: matches 1195. .1261 of consensus"
complement(4058. .4143)
/note="Dr000340 repeat: matches 133. .219 of consensus"
complement(4133. .4391)
/note="Dr000109 repeat: matches 6. .285 of consensus"
complement(4963. .25755)
/gene="SI:BY184L24.1"
complement(join(4963. .5283,6645. .6744,8032. .8164,
8355. .8405,11644. .11733,12218. .12346,19073. .19206,
19650. .19757,23460. .23813,25628. .25755))
/gene="SI:BY184L24.1"
/product="SI:BY184L24.1 (novel protein)"
/note="match: ESTs: Em:AI883018 Em:BI879331 Em:AI794629
Em:AI667648 Em:AA658628 Em:AI601737"
/evidence=not experimental
CDS complement(join(5208. .5283,6645. .6744,8032. .8164,
8355. .8405,11644. .11733,12218. .12346,19073. .19206,
19650. .19757,23460. .23685))
/gene="SI:BY184L24.1"
/note="match: proteins: Sw:AAH00645 Sw:Q92504 Tr:Q17996
Sw:Q9XUC4 Sw:Q31125 Tr:Q8WV10 Tr:Q96H72 Tr:AAL13757
Tr:Q9D1R4 Sw:Q9V3A4 Tr:Q9VDX1"
/codon_start=1
/evidence=not experimental
/product="SI:BY184L24.1 (novel protein)"
/protein_id="CAD58734.1"
/db_xref="GI:26788048"
/db_xref="GOA:Q8AW42"
/db_xref="UniProt/TREMBL:Q8AW42"
/translation="NM1QTAVQAQATAPAGPWSIKDLVDLOYLDLMSIDNLDVWF
CSLVGSIAGUSGIFPLVIPIEAGTALKTEAGCKRLKLLSFAIGGLGVFLHLLP
EAWATSPGSGHRYCYQGLWVIGLMSFLTEKMFPEVDGPDFTKTSFORTSSSS
DLSSQFSVSPQNGICSNNSDEKPTDISPYQPKTKISYGLNLLKNCIDNFTHGL
AVAGSFLVSRKVGFTLTFAILLIHPHEVGPFAILLRAGFORKWAARMQLSTALGGVL
GACFALCSQSOHGAENATTWILPFTSGGFLYIALVNVVDDLEETNPRNSLLQVLLF
SGIGWALLISAND"
5337. .5352
/note="3.2 copies 5 mer ATTGT 23% conserved"
complement(5377. .5478)
/note="ANGEL repeat: matches 192. .318 of consensus"
5520. .6524
/note="HARBINGER3_DR repeat: matches 2. .1064 of
consensus"
6528. .6583
/note="ANGEL repeat: matches 190. .245 of consensus"
6997. .7012
/note="2.3 copies 7 mer ATTTGTG 23% conserved"
7015. .7024
/note="10.0 copies 1 mer A 20% conserved"
7020. .7040
/note="3.0 copies 7 mer AAAATC 33% conserved"
```

```
repeat_region 7113. .7124
/note="3.0 copies 4 mer AAAT 24% conserved"
7323. .7386
/note="ANGEL repeat: matches 3. .70 of consensus"
7405. .7437
/note="16.5 copies 2 mer AT 66% conserved"
7438. .7573
/note="ANGEL repeat: matches 163. .307 of consensus"
7655. .7711
/note="8.1 copies 7 mer TATTTAT 30% conserved"
7673. .7706
/note="11.3 copies 3 mer TTA 27% conserved"
8051. .8066
/note="2.3 copies 7 mer AGAGCAA 32% conserved"
8421. .8450
/note="30.0 copies 1 mer A 24% conserved"
8424. .8441
/note="4.5 copies 4 mer AAAG 29% conserved"
8527. .8539
/note="2.6 copies 5 mer TAAAC 26% conserved"
8712. .8721
/note="2.5 copies 4 mer TATT 20% conserved"
8727. .8747
/note="5.2 copies 4 mer CAAA 24% conserved"
8731. .8760
/note="6.0 copies 5 mer AAAAC 44% conserved"
complement(8764. .9155)
/note="HARBINGER6_DR repeat: matches 1. .393 of consensus"
9164. .9233
/note="Dr000245 repeat: matches 355. .432 of consensus"
9237. .9378
/note="Dr000071 repeat: matches 64. .204 of consensus"
9468. .9492
/note="2.8 copies 9 mer GTTTAAAAAT 32% conserved"
10296. .10308
/note="13.0 copies 1 mer A 26% conserved"
10333. .11004
/note="TCIDR3 repeat: matches 2. .685 of consensus"
11004. .11219
/note="TCIDR3 repeat: matches 1001. .1225 of consensus"
11261. .11270
/note="2.5 copies 4 mer AATA 20% conserved"
11298. .11307
/note="10.0 copies 1 mer T 20% conserved"
11355. .11372
/note="2.0 copies 9 mer TAAATGAA 36% conserved"
11532. .11551
Query Match 50.8% Score 25.4; DB 5; Length 98204;
Best Local Similarity 74.4%; Pred. No. 72;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 5 CCCACCTGGATCTTGGCCCGCCCTTGAATTCACCAACCTC 47
DB 96932 CCCATCCGCATCTACCCCACTCATGATTCGCCCAACCTC 96890
RESULT 21
AC099796 132033 bp DNA linear PRI 28-DEC-2002
DEFINITION Homo sapiens chromosome 1 clone RP5-866120, complete sequence.
ACCESSION AC099796 AL162404
VERSION AC099796.2 GI:27413195
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132033)
AUTHORS Kaul, R. K., Olson, M. V., Zhou, Y., James, R. A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
Haugen, E. D.
TITLE Direct Submission
```

JOURNAL
REFERENCE

Unpublished
2 (bases 1 to 132033)
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
Direct Submission
Submitted (21-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

REFERENCE
AUTHORS

3 (bases 1 to 132033)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.

TITLE
JOURNAL

Direct Submission
Submitted (28-DEC-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Dec 28, 2002 this sequence version replaced gi:17027312.

COMMENT

----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgctg@u.washington.edu
Drafting Center: SC

----- Project Information

Center project name: chr-1
Center clone name: RP5-866L20 (sc0841)

----- Summary Statistics

Sequencing vector: plasmid; 58% of reads
Chemistry: Dye-terminator ET; 42% of reads
Chemistry: Dye-terminator Big Dye; 62% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 131632 bases at least Q40
Consensus quality: 131970 bases at least Q30
Consensus quality: 132027 bases at least Q20
Insert size: 132033; sum-of-contigs
Quality coverage: 8.4x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5': RP4-705F19 AL035415, 2428-bp overlap
3': RP11-240D10 AL590093, 2000-bp overlap
This is a partial submission, omitting 16999 bp of the 5' overlap.

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII EcoRI BglII

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

1123	1101	11398	11214	7830	7678
449	<800	2184	2165	5671	5644
512	<800	8065	8029	11476	11227
2814	2813	486	<800	31	<800
1247	1199	106	<800	9846	10010
4695	4665	2759	2797	2554	2491
3373	3396	691	<800	3354	3351
3371	3396	1119	1129	5799	5849
1726	1720	4445	4410	6667	6703
7553	7674	16360	16101	7445	7678
7059	7096	1176	1129	1444	1417
12237	11789	761	<800	955	981
1348	1329	3290	3383	227	<800
145	<800	18041	18453	1841	1843
15	<800	1543	1447	2713	2764
703	<800	6296	6386	9714	9566
4893	4872	3669	3631	4665	4634
978	987	734	704	7425	7396
1336	1329	699	<800	2633	2653
66	<800	3266	3236	473	<800
6554	6551	1459	1447	1794	1753
11681	11266	4639	4576	2158	2139
6563	6551	7026	7010	17231	17365
360	<800	16245	16101	992	981
6560	6551	3485	3530	2401	2491
8669	8488	4491	4576	868	849
2756	2813	1815	1817	3288	3351
3788	3739	1429	1447	51	<800
16627	16835	5139	5083	3152	3155
6300	6551	2827	2885	1691	1676
2340	2406	796	784	12159	11985
6932	7096	6000	6081	3608	3672
2550	2632	3696	3731	1313	1290
16563	16835	5245	5299	822	849
5528	5467	4264	4213	17305	17365
5648	5652	9418	9429	1908	1919

```
-----
FEATURES      source      Location/Qualifiers
1..132033
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RP5-866L20"
/clone_lib="RPC1 human PAC library 5"

ORIGIN
Query Match      50.8%; Score 25.4; DB 9; Length 132033;
Best Local Similarity 74.4%; Pred.No.70;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 5 CCCACCTCGATGCTTCCGCCGCCCTTGAAATTCACACCTTC 47
   || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 81768 CCATGCTGAATCCCTTCCTGCGCCCTCCATCTCCACCCCCC 81810

RESULT 22
AC025990/c
LOCUS      AC025990      166549 bp      DNA      linear      HTG 03-MAY-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-277A12 map 1, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
ACCESSION AC025990
VERSION AC025990.2 GI:7684488
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166549)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-277A12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166549)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Menues,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Scojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:7263208.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
-----
```

Center project name: L8019
Center clone name: 277 A.12
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148491 bases at least Q40
Consensus quality: 156797 bases at least Q30
Consensus quality: 160643 bases at least Q20
Insert size: 171000; agarose-fp
Insert size: 163649; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1034: contig of 1034 bp in length
* 1035 1134: gap of 100 bp
* 1135 2525: contig of 1391 bp in length
* 2526 2625: gap of 100 bp
* 2626 4050: contig of 1425 bp in length
* 4051 4150: gap of 100 bp
* 4151 5525: contig of 1375 bp in length
* 5526 5625: gap of 100 bp
* 5626 7121: contig of 1496 bp in length
* 7122 7221: gap of 100 bp
* 7222 8355: contig of 1134 bp in length
* 8356 8455: gap of 100 bp
* 8456 10273: contig of 1818 bp in length
* 10274 10373: gap of 100 bp
* 10374 12187: contig of 1814 bp in length
* 12188 12287: gap of 100 bp
* 12288 13713: contig of 1426 bp in length
* 13714 13813: gap of 100 bp
* 13814 16806: contig of 2993 bp in length
* 16807 16906: gap of 100 bp
* 16907 17920: contig of 1014 bp in length
* 17921 18020: gap of 100 bp
* 18021 21078: contig of 3058 bp in length
* 21079 21178: gap of 100 bp
* 21179 24583: contig of 3405 bp in length
* 24584 24683: gap of 100 bp
* 24684 29247: contig of 4564 bp in length
* 29248 29347: gap of 100 bp
* 29348 34482: contig of 5135 bp in length
* 34483 34582: gap of 100 bp
* 34583 37621: contig of 3039 bp in length
* 37622 37721: gap of 100 bp
* 37722 43206: contig of 5485 bp in length
* 43207 43306: gap of 100 bp
* 43307 45997: contig of 2691 bp in length
* 45998 46097: gap of 100 bp
* 46098 52449: contig of 6352 bp in length
* 52450 52549: gap of 100 bp
* 52550 57930: contig of 5381 bp in length
* 57931 58030: gap of 100 bp
* 58031 64917: contig of 6887 bp in length
* 64918 65017: gap of 100 bp
* 65018 74493: contig of 9476 bp in length
* 74494 74593: gap of 100 bp
* 74594 81921: contig of 7328 bp in length
* 81922 82021: gap of 100 bp
* 82022 89795: contig of 7774 bp in length
* 89796 89895: gap of 100 bp
* 89896 100646: contig of 10751 bp in length
* 100647 100746: gap of 100 bp
* 100747 109185: contig of 8439 bp in length

* 109186 109285: gap of 100 bp
 * 109286 120571: contig of 11286 bp in length
 * 120572 120671: gap of 100 bp
 * 120672 130560: contig of 9889 bp in length
 * 130561 130660: gap of 100 bp
 * 130661 145409: contig of 14749 bp in length
 * 145410 145509: gap of 100 bp
 * 145510 166549: contig of 21040 bp in length.

FEATURES

source

Location/Qualifiers
 1. .166549
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1"

/clone="RP11-277A12"
 /clone_lib="RPC1-11 Human Male BAC"

1. .1034

misc_feature /note="assembly_fragment"

misc_feature 1135. .2525 /note="assembly_fragment"

misc_feature 2626. .4050 /note="assembly_fragment"

misc_feature 4151. .5525 /note="assembly_fragment"

misc_feature 5626. .7121 /note="assembly_fragment"

misc_feature 7222. .8355 /note="assembly_fragment"

misc_feature 8456. .10273 /note="assembly_fragment"

misc_feature 10374. .12187 /note="assembly_fragment"

misc_feature 12288. .13713 /note="assembly_fragment"

misc_feature 13814. .16806 /note="assembly_fragment"

misc_feature 16907. .17920 /note="assembly_fragment"

misc_feature 18021. .21078 /note="assembly_fragment"

misc_feature 21179. .24583 /note="assembly_fragment"

misc_feature 24684. .29247 /note="assembly_fragment"

misc_feature 29348. .34482 /note="assembly_fragment"

misc_feature 34583. .37621 /note="assembly_fragment"

misc_feature 37722. .43206 /note="assembly_fragment"

misc_feature 43307. .45957 /note="assembly_fragment"

misc_feature 46098. .52449 /note="assembly_fragment"

misc_feature 52550. .57930 /note="assembly_fragment"

misc_feature 58031. .64917 /note="assembly_fragment"

misc_feature 65018. .74493 /note="assembly_fragment"

misc_feature 74594. .81921 /note="assembly_fragment"

misc_feature 82022. .89795 /note="assembly_fragment"

misc_feature 89896. .100646 /note="assembly_fragment"

misc_feature 100747. .109185 /note="assembly_fragment"

misc_feature 109286. .120571 /note="assembly_fragment"
 misc_feature 120672. .130560 /note="assembly_fragment"
 misc_feature 130661. .145409

Query Match 50.8%; Score 25.4; DB 2; Length 166549;

Best Local Similarity 74.4%; Pred. No. 69; Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 CCCACCCCTGGATCCTTGGCCCGCCCTTGAATTCACACCCCTC 47

DB 115182 CCATGCTGAATCCTTCCCTGCCCCCTCCACTCCACCCCCC 115140

RESULT 23

AC141729

LOCUS AC141729 201933 bp DNA linear HTG 19-MAR-2003
 DEFINITION Apis mellifera clone CH224-57G1, WORKING DRAFT SEQUENCE, 44
 unordered pieces.

ACCESSION AC141729

VERSION AC141729.1 GI:29123913

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.

SOURCE Apis mellifera (honey bee)

ORGANISM

Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
 Apidae; Apis.

REFERENCE 1 (bases 1 to 201933)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Albrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, B., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louieged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, S., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Umami, K., Vaquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 201933)

Worley, K.C.

repeat_region	204. .313	/note="5 copies 22 mer 71% conserved"	misc_feature	/note="2 copies 22 mer 97% conserved"
repeat_region	213. .312	/note="50 copies 2 mer at 75% conserved"	misc_feature	complement(66572. .67079)
repeat_region	215. .310		repeat_region	/note="match: GSS: Em:AQ233164"
repeat_region	313. .344	/note="8 copies 4 mer tata 87% conserved"	misc_feature	67080. .67189
misc_feature	complement(8801. .9503)		repeat_region	/note="match: GSS: Em:AQ000834"
repeat_region	9548. .9677	/note="65 copies 2 mer aa 60% conserved"	misc_feature	67194. .67226
misc_feature	complement(9959. .10571)		repeat_region	/note="11 copies 3 mer tgt 90% conserved"
misc_feature	/note="match: GSS: Em:A2516477"		misc_feature	complement(67439. .67670)
repeat_region	complement(12814. .13165)		repeat_region	/note="match: GSS: Em:AQ179536"
misc_feature	/note="match: GSS: Em:AQ333356"		misc_feature	70045. .70410
repeat_region	12985. .13158	/note="87 copies 2 mer aa 56% conserved"	repeat_region	/note="2 copies 183 mer 78% conserved"
misc_feature	13178. .13626		misc_feature	complement(70272. .70602)
repeat_region	match: GSS: Em:B82657"		repeat_region	/note="match: GSS: Em:AQ451491"
misc_feature	13467. .13556	/note="45 copies 2 mer aa 63% conserved"	repeat_region	70557. .70678
misc_feature	complement(19216. .19668)		repeat_region	/note="61 copies 2 mer aa 65% conserved"
misc_feature	/note="match: GSS: Em:AQ270125"		repeat_region	72107. .72150
misc_feature	/note="match: GSS: Em:AQ252652"		repeat_region	/note="2 copies 22 mer 97% conserved"
misc_feature	complement(19286. .19693)		repeat_region	72699. .72808
misc_feature	/note="match: GSS: Em:B57773"		repeat_region	/note="5 copies 22 mer 72% conserved"
misc_feature	/note="match: GSS: Em:B70777"		repeat_region	72708. .72799
misc_feature	complement(24855. .25328)		repeat_region	/note="23 copies 4 mer atat 72% conserved"
misc_feature	/note="match: GSS: Em:AQ269680"		repeat_region	72711. .72798
misc_feature	complement(24916. .25343)		repeat_region	/note="44 copies 2 mer ta 72% conserved"
misc_feature	/note="match: GSS: Em:B38750"		repeat_region	72810. .72865
misc_feature	complement(27844. .28327)		repeat_region	/note="14 copies 4 mer tata 75% conserved"
misc_feature	/note="match: GSS: Em:AQ359787"		repeat_region	72864. .72929
misc_feature	28046. .28551		misc_feature	/note="33 copies 2 mer at 66% conserved"
misc_feature	/note="match: GSS: Em:AQ702001"		misc_feature	75170. .75711
misc_feature	34614. .35016	/note="match: GSS: Em:B45119"	misc_feature	/note="match: GSS: Em:AQ885403"
misc_feature	complement(34943. .35437)		misc_feature	complement(78854. .79250)
misc_feature	/note="match: GSS: Em:AQ736443"		misc_feature	/note="match: GSS: Em:AQ583147"
misc_feature	complement(35334. .35685)		repeat_region	79388. .79580
misc_feature	/note="match: GSS: Em:AQ333372"		repeat_region	/note="match: GSS: Em:AQ383591"
repeat_region	complement(35730. .36160)		repeat_region	/note="match: GSS: Em:B65219"
repeat_region	36328. .36569	/note="121 copies 2 mer aa 57% conserved"	repeat_region	80751. .81478
repeat_region	36354. .36573		repeat_region	/note="match: GSS: Em:B54300"
misc_feature	39812. .40269	/note="10 copies 22 mer 60% conserved"	misc_feature	80778. .81231
misc_feature	match: GSS: Em:AQ285770"		repeat_region	/note="match: GSS: Em:AQ225169"
misc_feature	41208. .41586	/note="match: GSS: Em:AQ188545"	repeat_region	81829. .81894
misc_feature	complement(41289. .41670)		repeat_region	/note="3 copies 22 mer 84% conserved"
misc_feature	/note="match: GSS: Em:AQ443155"		repeat_region	81839. .81894
repeat_region	complement(41310. .41677)		repeat_region	/note="28 copies 2 mer gt 94% conserved"
repeat_region	42687. .42760	/note="match: GSS: Em:AQ108068"	repeat_region	81847. .82237
repeat_region	46370. .46419	/note="2 copies 37 mer 90% conserved"	repeat_region	/note="match: GSS: Em:AL232083"
misc_feature	complement(48477. .48591)		repeat_region	86384. .86389
repeat_region	49014. .49040	/note="9 copies 3 mer tta 92% conserved"	repeat_region	/note="13 copies 2 mer aa 92% conserved"
misc_feature	complement(53386. .53810)		repeat_region	93101. .93160
repeat_region	/note="match: GSS: Em:B45409"		repeat_region	/note="15 copies 4 mer agag 73% conserved"
repeat_region	55790. .55905		repeat_region	93102. .93159
repeat_region	55861. .55904	/note="29 copies 4 mer atat 69% conserved"	repeat_region	/note="29 copies 2 mer ga 74% conserved"
			repeat_region	93957. .94004
			repeat_region	/note="24 copies 2 mer ag 79% conserved"
			repeat_region	94125. .94290
			repeat_region	/note="83 copies 2 mer aa 56% conserved"
			repeat_region	96653. .97111
			repeat_region	/note="match: GSS: Em:AQ000813"
			repeat_region	100549. .100576
			repeat_region	/note="7 copies 4 mer gaat 92% conserved"
			repeat_region	101833. .102283
			repeat_region	/note="match: GSS: Em:AQ711871"
			repeat_region	103037. .103078
			repeat_region	/note="421 copies 2 mer tt 83% conserved"
			repeat_region	complement(103940. .104130)
			repeat_region	/note="match: GSS: Em:A2013061"
			repeat_region	complement(104324. .104827)
			repeat_region	/note="match: GSS: Em:AQ568181"
			repeat_region	complement(104684. .105209)
			repeat_region	/note="match: GSS: Em:AQ458439"
			repeat_region	complement(104709. .105187)
			repeat_region	/note="match: STS: Em:G51879"

Query Match

50.4%; Score 25.2; DB 9; Length 155358;

Best Local Similarity 71.7%; Pred. No. 82;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 ACCACCGCTGGATCCTTGTCCGCGCCCTTGAAATTCGAACCGCTCA 49 linear HTG 20-NOV-2002
|||||
DB 22632 ACCACCGCTGGTCCCTTCCCTGGTCTCTTTATTCTCAAGACTCA 22587

RESULT 31
AC134747

LOCUS 169164 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-354L3, WORKING DRAFT SEQUENCE.
ACCESSION AC134747
VERSION AC134747.2 GI:25139181
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1
Muzny,D,Marie, Metzker,M,Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D., Anylebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesat,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flegg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowic,C., Kraft,C.I., Lebow,H., Leván,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhsua,L., Loulesged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathewine,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemelema,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,Q., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steidle,M., Strong,R., Sutton,A., Svatek,A., Taborski,J., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlczky,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weitas,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

DIRECT SUBMISSION
Unpublished

TITLE
JOURNAL
PUBLISHED

REFERENCE 2 (bases 1 to 169164)

AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KBPL
Center clone name: CH230-354L3
----- Summary Statistics
Assembly program: Phrap; version 0.950329
Consensus quality: 155078 bases at least Q40
Consensus quality: 155942 bases at least Q30
Consensus quality: 156616 bases at least Q20
Estimated insert size: 158624; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as runs of N, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 169164: contig of 169164 bp in length.

Location/Qualifiers
1. .169164
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-354L3"
complement(11493..12348)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:B2253722"
complement(167376..168237)
/note="clone_boundary
clone_end:T7
site:
end_sequence:B2253720"

FEATURES
source
misc_feature
misc_feature

ORIGIN

Query Match 50.4%; Score 25.2; DB 2; Length 169164;
Best Local Similarity 71.7%; Pred. No. 82;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Query Match	50.4%	Score 25.2;	DB 2;	Length 169164;
Best Local Similarity	71.7%	Pred. NO. 82;		
Matches 33: Conservative	0:	Mismatches 13:	Indels 0:	Gaps 0:

JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

QY 1 CCTACCCACCTGGATCCTTGGCCGCCCGCCCTTGAATTCACCAACCT 46
|||||
DB 72075 CTTCCCATCCATGCTCTTCCGCGCCCTGACATTCCTACACT 72120
|||||

RESULT 32

AC095795

LOCUS

AC095795 188447 bp DNA linear HTG 10-SEP-2002
Rattus norvegicus clone CH230-9F18, *** SEQUENCING IN PROGRESS ***,
22 ordered pieces.

ACCESSION

AC095795.7 GI:22773036

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 188447)

Muzny,D,Marie, Metzker,M, Lee, A, Branzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Hernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kwis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshewa,L, Louleghed,H, Lozano,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhney,S, McLeod,M, McNeill,T, Meenen,B, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwackeleme,O, Okwuonu,G, Olarnpungsoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L-L, Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajd,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Wang,L, Wei,X, White,F, Williams,G, Willson,R, Wlarczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstock,G and Gibbs,R.A.

Direct Submission

Unpublished

REFERENCE

2 (bases 1 to 188447)

Worley,K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 188447)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 10, 2002 this sequence version replaced gi:22475295.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the

sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence

reads. Both end sequences and whole genome shotgun sequence only

contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDN

Center clone name: CH230-9F18

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 103178 bases at least Q40

Consensus quality: 109563 bases at least Q30

Consensus quality: 113768 bases at least Q20

Estimated insert size: 148240; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 22 contigs. Gaps between the contigs

are represented as runs of N. The order of the pieces

is believed to be correct as given, however the sizes

of the gaps between them are based on estimates that have

provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 12559: contig of 12559 bp in length

* 12560 12659: gap of unknown length

* 12660 52547: contig of 39888 bp in length

* 52548 52648: gap of unknown length

* 52648 59120: contig of 6473 bp in length

* 59121 59220: gap of unknown length

* 59221 159636: contig of 100416 bp in length

* 159637 159736: gap of unknown length

* 159737 160752: contig of 1016 bp in length

* 160753 160852: gap of unknown length

* 160853 161861: contig of 1009 bp in length

* 161862 161961: gap of unknown length

* 161962 163112: contig of 1151 bp in length

* 163113 163212: gap of unknown length

* 163213 164581: contig of 1369 bp in length

* 164582 164681: gap of unknown length

* 164682 166174: contig of 1493 bp in length

* 166175 166274: gap of unknown length

* 166275 167719: contig of 1445 bp in length

* 167720 167819: gap of unknown length

* 167820 168826: contig of 1007 bp in length

* 168827 168926: gap of unknown length

* 168927 170062: contig of 1136 bp in length

* 170063 170162: gap of unknown length

* 170163 171637: contig of 1475 bp in length

* 171638 171737: gap of unknown length

* 171738 171739: contig of 1454 bp in length

* 171740 173291: gap of unknown length

* 173292 174474: contig of 1183 bp in length

* 174475 174474: contig of 1183 bp in length

* 174475 174474: contig of 1183 bp in length

TITLE

JOURNAL

AUTHORS

TITLE

```

* 174475 174574: gap of unknown length
* 174575 175801: contig of 1227 bp in length
* 175802 175901: gap of unknown length
* 175902 177378: contig of 1477 bp in length
* 177379 177478: gap of unknown length
* 177479 179074: contig of 1596 bp in length
* 179075 179174: gap of unknown length
* 179175 181077: contig of 1903 bp in length
* 181078 181177: gap of unknown length
* 181178 182200: contig of 1023 bp in length
* 182201 182300: gap of unknown length
* 182301 186090: contig of 3790 bp in length
* 186091 186191: gap of unknown length
* 186191 188447: contig of 2257 bp in length.

FEATURES
  source
    1..188447
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clone="CH230-9F18"
      /notes="wgs contig"
      /note="wgs_contig"

  misc_feature
    52648..55815
      /note="wgs contig"

  misc_feature
    76145..77242
      /note="wgs_contig"

ORIGIN
Query Match      50.4%; Score 25.2; DB 2; Length 188447;
Best Local Similarity 71.7%; Pred. No. 81;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCTTGGCCCCCGCCCTTGAATTCACACCT 46
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 114380 CCTCCCATCCATGCTCTTCCGCCGCCCTGACATTCCTACACT 114425

RESULT 33
AC009959
LOCUS      Homo sapiens chromosome 14 clone RP11-353G11, WORKING DRAFT
DEFINITION
SEQUENCE   3 unordered pieces.
ACCESSION  AC009959
VERSION    AC009959.6 GI:8748894
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 190177)
           Waterston,R.H.
           The sequence of Homo sapiens clone
           Unpublished
REFERENCE  2 (bases 1 to 190177)
           Waterston,R.H.
           Direct Submission
           Submitted (08-SEP-1999) Genome Sequencing Center, Washington
           University School of Medicine, 444 Forest Park Parkway, St. Louis,
           MO 63108, USA
COMMENT    On Jun 27, 2000 this sequence version replaced gi:85669954.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_NH0353G11
----- Project Information -----
Sequencing vector: p1asmid; 12k
Chemistry: Dye-terminator Big Dye; 12k of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187629 bases at least Q40
Consensus quality: 188853 bases at least Q30

```

```

Consensus quality: 189456 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 189977; sum-of-contigs
Quality coverage: 5.10 in Q20 bases; agarose-fp
Quality coverage: 5.42 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 20606: contig of 20606 bp in length
* 20607 20706: gap of unknown length
* 20707 91735: contig of 71029 bp in length
* 91736 91835: gap of unknown length
* 91836 190177: contig of 98342 bp in length.

FEATURES
  source
    1..190177
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="14"
      /clone="RP11-353G11"
      /note="assembly_name:Contig11"
      /note="assembly_name:Contig12"
      /note="assembly_name:Contig13"

  misc_feature
    1..20606
      /clone_end:SP6
      /vector_side:left

  misc_feature
    20707..91735
      /clone_end:T7
      /vector_side:left

  misc_feature
    91836..190177
      /note="assembly_name:Contig13"

ORIGIN
Query Match      50.4%; Score 25.2; DB 2; Length 190177;
Best Local Similarity 78.9%; Pred. No. 81;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGATCTTGGCCGCCCGCCCTTGAATTCACACCTC 47
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94513 CCTGTGCTGACCCCTGTCTTGCATTCACCAACCTC 94550

RESULT 34
AL929424/c
LOCUS      Mouse DNA sequence from clone RP23-24F19 on chromosome 2, complete
DEFINITION
ACCESSION  AL929424
VERSION    AL929424.7 GI:27657509
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 204020)
           Heath,P.
           Direct Submission
           Submitted (01-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           humquerry@sanger.ac.uk
           Clone requests: clonerequest@sanger.ac.uk
           On Jan 12, 2003 this sequence version replaced gi:27111676.
           Sequence from the Mouse Genome Sequencing Consortium whole genome
           shotgun may have been used to confirm this sequence. Sequence data
           from the whole genome shotgun alone has only been used where it has
           a phred quality of at least 30.
           ----- Genome Center -----
           Center: Wellcome Trust Sanger Institute
           Center code: SC

```



```

Query Match      50.4%; Score 25.2; DB 10; Length 209383;
Best Local Similarity 71.7%; Pred. No. 80;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 ACCACCTGATCTTGGCCCGCCCTTGATTCCTCAACCTCCCA 49
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125294 ACCACCTGATCTTGGCCCGCCCGCCGAGAACTCAACCTCCAA 125339

AC146886      210493 bp      DNA      linear      HTG 16-OCT-2003
Callithrix jacchus clone CH259-59B13, WORKING DRAFT SEQUENCE, 5
unordered pieces.

AC146886      GI:37693549
HTG: HTGS PHASE1; HTGS DRAFT.
KEYWORDS
SOURCE
ORGANISM
Callithrix jacchus (white-tufted-ear marmoset)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
Callithrix.
1 (bases 1 to 210493)
Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R.,
Karlsins,E., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
Green,E.D.

NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 210493)
Green,E.D.
Direct Submission
Submitted (16-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@hgrl.nih.gov
----- Project Information
Center project name: ftc
Center clone name: 059B13
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 209180 bases at least Q40
Consensus quality: 209325 bases at least Q30
Consensus quality: 209367 bases at least Q20
Insert size: 213000; agarose-fp
Insert coverage: 10.12x in Q20 bases;
Quality coverage: 10.26x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2669: contig of 2669 bp in length
* 2670 2769: gap of unknown length
* 2770 11747: contig of 8978 bp in length

```

```

FEATURES
source
1..210493
/organism="Callithrix jacchus"
/mol_type="genomic DNA"
/db_xref="taxon:9483"
/clone="CH259-59B13"
/clone_lib="CH259"
1..2669
/note="assembly_fragment"
2770..11747
/note="assembly_fragment"
11848..23051
/note="assembly_fragment"
23152..99160
/note="assembly_fragment"
99261..210493
/note="assembly_fragment"
clone_end:SP6
vector_side:right

```

```

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

ORIGIN

```

Query Match      50.4%; Score 25.2; DB 2; Length 210493;
Best Local Similarity 71.7%; Pred. No. 80;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCTTGGCCCGCCCTTGATTCCTCAACCTCC 47
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95884 CTTCCTCCCTCTTACCCCATGTATCTCCCATGATTCGCAACCTC 95929

```

RESULT 37

```

AC146883
LOCUS
DEFINITION
Callithrix jacchus clone CH259-225J7, WORKING DRAFT SEQUENCE, 4
ordered pieces.
ACCESSION
AC146883.2 GI:38093728
VERSION
HTG: HTGS PHASE2; HTGS DRAFT.
KEYWORDS
Callithrix jacchus (white-tufted-ear marmoset)
SOURCE
Callithrix jacchus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
Callithrix.

```

REFERENCE

```

1 (bases 1 to 216177)
Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R.,
Karlsins,E., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 216177)
Green,E.D.
Direct Submission
Submitted (16-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 216177)
Green,E.D.
Direct Submission
Submitted (31-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

```

TITLE

```

JOURNAL
REFERENCE
2 (bases 1 to 216177)
AUTHORS
Green,E.D.
TITLE
JOURNAL
Submitted (16-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE
3 (bases 1 to 216177)
AUTHORS
Green,E.D.
TITLE
JOURNAL
Submitted (31-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

```

COMMENT

On Oct 31, 2003 this sequence version replaced gi:37693546.

----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@nhgri.nih.gov
 ----- Project Information
 Center project name: ftb
 Center clone name: 225707

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 215611 bases at least Q40
 Consensus quality: 215808 bases at least Q30
 Consensus quality: 215869 bases at least Q20
 Insert size: 216000; agarose-fp
 Insert size: 215877; sum-of-contigs
 Quality coverage: 11.46x in Q20 bases; agarose-fp
 Quality coverage: 11.46x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 31979: contig of 31979 bp in length
 * 31980 32079: gap of unknown length
 * 32080 159004: contig of 126925 bp in length
 * 159005 159104: gap of unknown length
 * 159105 193998: contig of 34894 bp in length
 * 193999 194098: gap of unknown length
 * 194099 216177: contig of 22079 bp in length.

FEATURES

source

Location/Qualifiers
 1..216177
 /organism="Callithrix jacchus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9483"
 /clone="CH259-225J7"
 /clone_lib="CH259"

misc_feature

1..32591
 /note="clone overlaps with GenBank Accession Number AC146661 clone CH259-140516 (center project name fta)"

misc_feature

1..31979
 /note="assembly_fragment"

clone_end:17
 vector_side:left

misc_feature

32080..159004
 /note="assembly_fragment"

misc_feature

159105..193998
 /note="assembly_fragment"

misc_feature

194099..216177
 /note="assembly_fragment"

clone_end:SP6
 vector_side:right

misc_feature

194944..216177

ORIGIN

Query Match 50.4%; Score 25.2; DB 2; Length 216177;
 Best Local Similarity 71.7%; Pred. No. 80;
 Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 2 CTACCCACCCTGGATCTTGCCCGCCCTTGGAATTCACCAACCTC 47
 Db 214664 CTTCCCTCTCTTACCCATGATCTCCCATGATTCACCAACCTC 214709

RESULT 38

BX294436

LOCUS

BX294436 218761 bp DNA linear HTG 02-APR-2003
 MUS musculus chromosome 2 clone RP23-118A19, *** SEQUENCING IN
 PROGRESS ***, 5 unordered pieces.

ACCESSION

BX294436

VERSION

BX294436.2 GI:29500911

KEYWORDS

HTG; HTGS PHASE1

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 218761)

Sims, S.

Direct Submission

Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 2, 2003 this sequence version replaced gi:29125447.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BM118A19
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 217328 bases at least Q40
 Consensus quality: 217830 bases at least Q30
 Consensus quality: 218153 bases at least Q20
 Insert size: 218361; sum-of-contigs
 Insert size: 171487; 1.8% error; agarose-fp
 Quality coverage: 6.27x in Q20 bases; sum-of-contigs Quality
 coverage: 8.13x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 15344: contig of 15344 bp in length
 * 15345 15444: gap of 100 bp
 * 15445 97450: contig of 82006 bp in length
 * 97451 97550: gap of 100 bp
 * 97551 138871: contig of 41321 bp in length
 * 138872 138971: gap of 100 bp
 * 138972 177076: contig of 38105 bp in length
 * 177077 177176: gap of 100 bp
 * 177177 218761: contig of 41585 bp in length.

Location/Qualifiers
 1..218761

FEATURES
 source

1..218761


```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-118A19"
/clone_lib="RPCI-23"
1. .15344
/notes="assembly_fragment:00362
fragment_chain:1"
misc_feature
15445. .97450
/notes="assembly_fragment:01130
fragment_chain:1"
misc_feature
97551. .138871
/notes="assembly_fragment:01492
fragment_chain:1"
misc_feature
138972. .177076
/notes="assembly_fragment:00001"
misc_feature
177177. .218761
/notes="assembly_fragment:00886.0"

ORIGIN
Query Match 50.4%; Score 25.2; DB 2; Length 218761;
Best Local Similarity 71.7%; Pred. No. 80;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 5 CCACCGTCGATCCTTGCCCGCCCTTGAAATTCACACCTCCAC 50
DB 151 CCCTCCCTCTTCCCTTCCCTCCCTCCCATCTTTCTCACCCCTCCAC 196

RESULT 39
AC146674
LOCUS
DEFINITION
AC146674 228016 bp DNA linear HTG 09-OCT-2003
Callithrix jacchus clone CH259-338M13, WORKING DRAFT SEQUENCE, 7
ordered pieces.
ACCESSION
AC146674.2 GI:37591190
VERSION
HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS
Callithrix jacchus (white-tufted-ear marmoset)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
Callithrix.
1 (bases 1 to 228016)
Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R.,
Karling,E., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
Maskeri,B., McDowell,J., Mullikin,J.C., Paquirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
Green,E.D.
NISC Comparative Sequencing Initiative
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 228016)
Green,E.D.
AUTHORS
Direct Submission
TITLE
Submitted (16-SEP-2003) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
JOURNAL
3 (bases 1 to 228016)
Green,E.D.
AUTHORS
Direct Submission
TITLE
Submitted (09-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
JOURNAL
On Oct 9, 2003 this sequence version replaced gi:34740347.
COMMENT
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov

```

```

----- Project Information
Center project name: ftd
Center clone name: 338M13

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

```

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 226812 bases at least Q40
Consensus quality: 227090 bases at least Q30
Consensus quality: 227338 bases at least Q20
Insert size: 222000; agarose-fp
Insert size: 227416; sum-of-contigs
Quality coverage: 11.93x in Q20 bases; agarose-fp
Quality coverage: 11.64x in Q20 bases; sum-of-contigs

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 34176: contig of 34176 bp in length
* 34177 34276: gap of unknown length
* 34277 43742: contig of 9466 bp in length
* 43743 43842: gap of unknown length
* 43843 118187: contig of 74345 bp in length
* 118188 118287: gap of unknown length
* 118288 143028: contig of 24741 bp in length
* 143029 143129: gap of unknown length
* 143129 207941: contig of 64813 bp in length
* 207942 215832: contig of 7791 bp in length
* 215833 215932: gap of unknown length
* 215933 228016: contig of 12084 bp in length.
FEATURES
Location/Qualifiers
1. 228016
/organism="Callithrix jacchus"
/mol_type="genomic DNA"
/db_xref="taxon:9483"
/clone="CH259-338M13"
/clone_lib="CH259"
1. 34176
/notes="assembly_fragment
clone_end:17
vector_side:left"
misc_feature
34277. .43742
/notes="assembly_fragment"
misc_feature
43843. .118187
/notes="assembly_fragment"
misc_feature
118288. .143028
/notes="assembly_fragment"
misc_feature
143129. .207941
/notes="assembly_fragment"
misc_feature
208042. .215832
/notes="assembly_fragment"
misc_feature
215933. .228016
/notes="assembly_fragment"

```

```

clone end:SP6
vector_side:right"

ORIGIN
Query Match      50.4%; Score 25.2; DB 2; Length 228016;
Best Local Similarity 71.7%; Pred. No. 79;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGCCTCCCGCCCTTGAATTCCTCAACCCCTC 47
DB 19725 CTTCCCTCCCTTTACCCCATGTATCTCCCATGAATTCGCAACCCCTC 19770

RESULT 40
AC140217/c
LOCUS AC140217 222953 bp DNA linear HTG 29-JUL-2004
DEFINITION Mus musculus chromosome 8 clone RP23-27905, WORKING DRAFT SEQUENCE,
8 unordered pieces.
ACCESSION AC140217.2 GI:50811808
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 232953)
AUTHORS Wilson,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 232953)
AUTHORS McPherson,J.D. and Waterston,R.H..
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 232953)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 29, 2004 this sequence version replaced gi:28475454.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BA0279005
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 229691 bases at least Q40
Consensus quality: 230683 bases at least Q30
Consensus quality: 231142 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1377: contig of 1377 bp in length
* 1378 1477: gap of unknown length
* 1478 2701: contig of 1224 bp in length
* 2702 2801: gap of unknown length
* 2802 4003: contig of 1202 bp in length
* 4004 4103: gap of unknown length

us-10-010-476-13.rge
5548: contig of 1445 bp in length
5648: gap of unknown length
47235: contig of 41587 bp in length
47236: gap of unknown length
11198: contig of 63863 bp in length
11199: gap of unknown length
11299: contig of 120273 bp in length
231572: gap of unknown length
231672: contig of 1282 bp in length.
Location/Qualifiers
1. 232953
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/clone="RP23-27905"
1. 1377
/note="assembly_name:Contig28"
1478..2701
/note="assembly_name:Contig33"
2802..4003
/note="assembly_name:Contig34"
4104..5548
/note="assembly_name:Contig44"
5649..47235
/note="assembly_name:Contig45"
47336..11198
/note="assembly_name:Contig46"
111299..231571
/note="assembly_name:Contig47"
231672..232953
/note="assembly_name:Contig6"

ORIGIN
Query Match 50.4%; Score 25.2; DB 2; Length 232953;
Best Local Similarity 71.7%; Pred. No. 79;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 ACCCACCTGGATCCTTGCCTCCCGCCCTTGAATTCCTCAACCCCTCCA 49
DB 51856 ACCCACCTGGATCCTTGCCTCCCGCCCGCCGAGAACTCACAACCCCTCAA 51811

RESULT 41
AC127753
LOCUS AC127753 237359 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-21M19, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
ACCESSION AC127753
VERSION AC127753.3 GI:25079334
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 237359)
AUTHORS Murry,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allien,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelam, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Paesternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmami, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 237359)
 Worley, K.C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 237359)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23195188.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GZRT
 Center clone name: CH230-21M19
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 186995 bases at least Q40
 Consensus quality: 190628 bases at least Q30
 Consensus quality: 193100 bases at least Q20

Estimated insert size: 198498; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 6 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 51279: contig of 51279 bp in length
 * 51280 51379: gap of unknown length
 * 51380 153527: contig of 102148 bp in length
 * 153528 153627: gap of unknown length
 * 153628 179889: contig of 26262 bp in length
 * 179890 179889: gap of unknown length
 * 179990 212085: contig of 32096 bp in length
 * 212086 212185: gap of unknown length
 * 212186 235740: contig of 23555 bp in length
 * 235741 235840: gap of unknown length
 * 235841 237359: contig of 1519 bp in length.
 Location/Qualifiers
 1..237359
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-21M19"
 /complement(22428..23293)
 /note="clone boundary
 clone end: T7
 site:
 end_sequence: BH272355"
 misc_feature
 ORIGIN
 Query Match 50.4%; Score 25.2; DB 2; Length 237359;
 Best Local Similarity 78.9%; Pred. No. 79;
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 CCTACCCACCTGGATCTTGGCCCGCCCTTGATTC 38
 |||||
 Db 157142 CCTATCGTCCTGGATCTTGGCCCGCCCTTGATTC 157179
 RESULT 42
 ACL07344
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-117F21, *** SEQUENCING IN PROGRESS
 ***, 5 unordered pieces.
 AC107344 263882 bp DNA linear HTG 22-SEP-2002
 Rattus norvegicus
 AC107344 GI:23195787
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 263882)
 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

```

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Garcia,A., Garner,T., Garza,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Foster,P.,
Gregoire,B., Geis,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebirt,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhea,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwar,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Mitosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleleh,O., Okwuonu,G., Olarnpunagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,R., Warren,X., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 263882)
Worley,K.C.
Direct Submission
Submitted (19-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 263882)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21737124.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GOOX
Center clone name: CH230-117F21
-----
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226341 bases at least Q40
Consensus quality: 229458 bases at least Q30
Consensus quality: 231425 bases at least Q20
Estimated insert size: 248600; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 27027: contig of 27027 bp in length
* 27028 27127: gap of unknown length
* 27128 254735: contig of 227608 bp in length
* 254736 254835: gap of unknown length
* 254836 256430: contig of 1595 bp in length
* 256431 256530: gap of unknown length
* 256531 258485: contig of 1955 bp in length
* 258486 263882: gap of unknown length
* 263883 263882: contig of 5297 bp in length.
FEATURES
Location/Qualifiers
source
1..263882
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-117F21"
misc_feature
1..1357
/note="wgs_end_extension
clone_end:T7"
5003..5484
/note="clone_boundary
clone_end:T7"
site:EcoRI
end_sequence:BH280995"
27128..31894
/note="wgs_contig"
ORIGIN
Query Match 50.4%; Score 25.2; DB 2; Length 263882;
Best Local Similarity 78.9%; Pred.No.78;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 CCTATCCACCTGGATCCTTGGCCGCCCTTGAATTC 38
Db 103091 CCTATCGGCTGGATCCTTGGCCGCCCTTGAATTC 103128
RESULT 43
AC015310/c
LOCUS AC015310 53950 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC015310
VERSION AC015310.1 GI:6436025
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 53950)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211259 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..53950
/organism="Drosophila melanogaster"
FEATURES
source

```

```

ORIGIN
Query Match          50.0%; Score 25; DB 9; Length 101445;
Best Local Similarity 69.4%; Pred. No. 1e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 CTATCCACCCCTGATCCTTGGCCCGCCCTTGGAATTCCCAACCCCTCAC 50
   |||||||
Db 39280 CTCCACACCCCTGGGATCATCTCCAGCACCCTCAATTCCCTATATGTCAC 39232
   |||||||

RESULT 45
HS998C11
LOCUS
DEFINITION
Human DNA sequence from clone RP5-998C11 on chromosome
20q13.13-13.2 Contains part of the KIAA1415 gene, ESTs, STSs and
GSSs, complete sequence.
ACCESSION
AL035106
VERSION
AL035106.28 GI:11323314
KEYWORDS
HTG; KIAA1415.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
Bates,K.
Direct Submission
Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: hunquery@sanger.ac.uk Clone
requests: clonesrequest@sanger.ac.uk
COMMENT
On Nov 24, 2000 this sequence version replaced gi:10198623.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
This sequence is the entire insert of clone RP5-998C11 The true
left end of clone RP5-998C11 is at 114259 in this sequence. The true
right end of clone RP11-263H4 is at 60004 in this sequence. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP5-998C11 is from
the library RPCI-5 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
FEATURES
Location/Qualifiers
1..131007
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.13-13.2"
/clone="RP5-998C11"
/clone_lib="RPCI-5"
repeat_region 32..328
/note="AluDb repeat: matches 1..296 of consensus"
repeat_region 425..639
/note="MER20 repeat: matches 3..218 of consensus"

```


Db 52075 CTTCTCCCTCCCTCTCTTACCATTCATTCCATCCCAACCCCA 52123
||||| ||||| ||||| || || || || ||||| || |||||

RESULT 46
AC104506/c
LOCUS AC104506 131769 bp DNA linear PRI 20-AUG-2002
DEFINITION Homo sapiens chromosome 1 clone RP4-549L20, complete sequence.
ACCESSION AC104506 AL0096820
VERSION AC104506.2 GI:22325040
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 131769)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
Haugen, E.D.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131769)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 131769)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
Haugen, E.D.

TITLE Direct Submission
JOURNAL Submitted (20-AUG-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Aug 20, 2002 this sequence version replaced gi:17544718.

----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP4-549L20 (sc0509)
----- Summary Statistics
Sequencing vector: plasmid; 37% of reads
Chemistry: Dye-terminator ET; 68% of reads
Chemistry: Dye-terminator Big Dye; 32% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 131593 bases at least Q40
Consensus quality: 131762 bases at least Q30
Consensus quality: 131769 bases at least Q20
Insert size: 131769; sum-of-contigs
Quality coverage: 7.2x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': Mapping in progress
3': RP11-86R24 (UWGC:sc0310) AC093428, 31725-bp overlap

----- Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing

problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI				BglII				HindIII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8022	8210	8063	8058	7708	7683						
-----	-----	-----	-----	-----	-----						
2184	2194	5671	5763	449	<800						
-----	-----	-----	-----	-----	-----						
8065	8210	4349	4453	512	<800						
-----	-----	-----	-----	-----	-----						
486	<800	2271	2252	2814	2853						
-----	-----	-----	-----	-----	-----						
106	<800	160	<800	1247	1215						
-----	-----	-----	-----	-----	-----						
4349	4406	10001	9974	4695	4738						
-----	-----	-----	-----	-----	-----						
16557	16279	2022	2012	3373	3379						
-----	-----	-----	-----	-----	-----						
3347	3371	5154	5136	8182	8271						
-----	-----	-----	-----	-----	-----						
5958	5959	740	773	7073	7106						
-----	-----	-----	-----	-----	-----						
5612	5589	1870	1850	267	<800						
-----	-----	-----	-----	-----	-----						
1018	1011	2152	2252	809	841						
-----	-----	-----	-----	-----	-----						
2500	2543	1240	1227	3171	3139						
-----	-----	-----	-----	-----	-----						
1867	1863	783	773	227	<800						
-----	-----	-----	-----	-----	-----						
4015	4011	8398	8435	1769	1787						
-----	-----	-----	-----	-----	-----						
1146	1132	2701	2700	543	<800						
-----	-----	-----	-----	-----	-----						
13001	12580	2050	2151	537	<800						
-----	-----	-----	-----	-----	-----						
1653	1628	2199	2252	1646	1613						
-----	-----	-----	-----	-----	-----						
900	898	632	<800	3212	3139						
-----	-----	-----	-----	-----	-----						
4904	4865	38	<800	4770	4738						
-----	-----	-----	-----	-----	-----						
8215	8210	150	<800	834	841						
-----	-----	-----	-----	-----	-----						
865	898	2001	2012	46	<800						
-----	-----	-----	-----	-----	-----						
1995	1954	2036	2012	2736	2758						
-----	-----	-----	-----	-----	-----						
1466	1440	5611	5510	4792	4738						
-----	-----	-----	-----	-----	-----						
1298	1279	4482	4453	1406	1398						
-----	-----	-----	-----	-----	-----						
2090	2107	8654	8739	849	841						
-----	-----	-----	-----	-----	-----						
10815	10639	3545	3524	1811	1787						
-----	-----	-----	-----	-----	-----						

4697	4653	41	<800	1040	1037
118	<800	252	<800	7712	7683
3839	3793	1477	1455	414	<800
20391	20586	527	<800	2284	2306
5575	5589	5712	5763	4366	4354
725	738	1346	1339	4974	4917
		4058	4036	1210	1215
		1237	1227	6699	6687
		3197	3210	2158	2217
		1270	1227	9761	9782
		1828	1850	3425	3379
		925	927	1625	1613
		3709	3684	181	<800
		5869	5763	291	<800
		1567	1518	2722	2758
		1040	1048	1449	1398
		1350	1339	1852	1787
		2865	2846	3089	3139
		639	<800	690	<800
		122	<800	5952	5870
		2950	2971	1099	1103
		5419	5317	2059	2036
		4170	4156	6930	6909
		9236	9209	4807	4738
				954	956
				1768	1787
				1677	1713
				457	<800
				656	<800

FEATURES	Location/Qualifiers
source	1. .131769
Query Match	50.0%; Score 25; DB 9; Length 131769;
Best Local Similarity	69.4%; Pred.No. 99;
Matches 34; Conservative	0; Mismatches 15; Indels 0; Gaps 0;
Qy	2 CTACCCACCGCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
Db	101779 CTATCATCCTCGAGATTTCCTCATGCCCTTTCATACCAACCCCCAC 101731

RESULT 47	
AC093428	
LOCUS	AC093428 152858 bp DNA linear PRI 15-JAN-2002

Homo sapiens chromosome 1 clone RP11-86F24, complete sequence.
AC093428 AL157715
AC093428.2 GI:18151023
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152858)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
Unpublished
2 (bases 1 to 152858)
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and Haugen, E.D.
Direct Submission
Submitted (23-AUG-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 152858)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
Direct Submission
Submitted (15-JAN-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
On Jan 15, 2002 this sequence version replaced gi:15281279.

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgctgs@u.washington.edu
Drafting Center: SC

Project Information
Center project name: chr-1
Center clone name: RP11-86F24 (sc0310)

Summary Statistics
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET; 89% of reads
Chemistry: Dye-terminator Big Dye; 11% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 152817 bases at least Q40
Consensus quality: 152857 bases at least Q30
Consensus quality: 152858 bases at least Q20
Insert size: 152863; sum-of-contigs
Quality coverage: 9.6x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': Mapping in progress
3': RP4-549L20 (UMGC:sc0509) AC104506

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and

vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI				BglII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----
8696	8699	2327	2299	2688	2703		
-----	-----	-----	-----	-----	-----	-----	-----
6	<800	6382	6605	2067	2086		
-----	-----	-----	-----	-----	-----	-----	-----
20385	20454	512	<800	6161	6365		
-----	-----	-----	-----	-----	-----	-----	-----
5572	5470	449	<800	1567	1532		
-----	-----	-----	-----	-----	-----	-----	-----
725	733	5631	5571	1034	1058		
-----	-----	-----	-----	-----	-----	-----	-----
6495	6445	1099	1163	1350	1323		
-----	-----	-----	-----	-----	-----	-----	-----
12117	11799	2059	2038	2865	2918		
-----	-----	-----	-----	-----	-----	-----	-----
110	<800	6930	6988	639	<800		
-----	-----	-----	-----	-----	-----	-----	-----
729	733	4807	4750	122	<800		
-----	-----	-----	-----	-----	-----	-----	-----
3053	3082	954	972	2950	2918		
-----	-----	-----	-----	-----	-----	-----	-----
3426	3445	1765	1685	5419	5357		
-----	-----	-----	-----	-----	-----	-----	-----
723	733	1677	1685	4167	4184		
-----	-----	-----	-----	-----	-----	-----	-----
1346	1331	457	<800	9236	9189		
-----	-----	-----	-----	-----	-----	-----	-----
279	<800	656	<800	1319	1323		
-----	-----	-----	-----	-----	-----	-----	-----
73	<800	7465	7465	1081	1058		
-----	-----	-----	-----	-----	-----	-----	-----
6911	6951	3401	3349	939	952		
-----	-----	-----	-----	-----	-----	-----	-----
4738	4710	11734	11475	5431	5173		
-----	-----	-----	-----	-----	-----	-----	-----
3950	3912	554	<800	6045	5989		
-----	-----	-----	-----	-----	-----	-----	-----
5904	5876	295	<800	750	769		
-----	-----	-----	-----	-----	-----	-----	-----
139	<800	6551	6605	5223	5357		
-----	-----	-----	-----	-----	-----	-----	-----
5451	5470	2337	2195	555	<800		
-----	-----	-----	-----	-----	-----	-----	-----
229	<800	134	<800	6054	5989		
-----	-----	-----	-----	-----	-----	-----	-----
158	<800	18841	18893	351	<800		
-----	-----	-----	-----	-----	-----	-----	-----
3108	3082	2098	2038	1877	1913		
-----	-----	-----	-----	-----	-----	-----	-----
1396	1331	7084	6988	2015	2086		
-----	-----	-----	-----	-----	-----	-----	-----
1910	1895	1287	1355	2115	2086		
-----	-----	-----	-----	-----	-----	-----	-----
1342	1331	5704	5571	4827	4788		
-----	-----	-----	-----	-----	-----	-----	-----
483	<800	5435	5571	871	883		
-----	-----	-----	-----	-----	-----	-----	-----
5436	5470	2205	2299	7700	7740		
-----	-----	-----	-----	-----	-----	-----	-----
1806	1895	1392	1355	957	952		
-----	-----	-----	-----	-----	-----	-----	-----
931	934	5961	6036	1249	1244		

FEATURES

source
Location/Qualifiers
1..152858
/organism="Homo sapiens"
/mol_type="genomic DNA"

Query Match 50.0%; Score 25; DB 9; Length 152858;
Best Local Similarity 69.4%; Pred. No. 98;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCACCTCCAC 50
Db 151124 CTATCATCCTGGAGATTCTCTCATGCGCTTTCATACCCACCTCCAC 151172

RESULT 48

AC010708/c
LOCUS AC010708 188766 bp DNA linear INV 17-MAR-2001
DEFINITION Drosophila melanogaster, chromosome X, region 17A-17B, BAC clone
BACR10E05, complete sequence.
ACCESSION AC010708
VERSION AC010708.10 GI:13374649
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 188766)

REFERENCE
AUTHORS
Celniker, S.E., Adams, J.M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banton, J., Beeson, K.Y., Buesam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dreeneke, D., Farfan, D., Ferreira, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunco, J., Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome X, region 17A-17B

Unpublished

2 (bases 1 to 188766)

REFERENCE
AUTHORS
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E., Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission

Submitted (20-SEP-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Mar 17, 2001 this sequence version replaced gi:6978366.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu.

FEATURES
source
1..188766
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strains="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="17A-17B"
/clone="BACR10805 (D901)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACs3.6)"

ORIGIN

Query Match 50.0%; Score 25; DB 3; Length 188766;
Best Local Similarity 75.6%; Pred. No. 96;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 ACCACCCCTGGATCTTGGCCCGCCCTTGAATTCACACC 44
|||||

Db 71953 ACCACCCCGCGGTACCCCGCCCTCGTATTCGACC 71913
|||||

RESULT 49
AC007207
LOCUS
DEFINITION
AC007207 Homo sapiens 12 BAC RP11-320N7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
AC007207
VERSION
AC007207.22 GI:6466489

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 191877)
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bouda, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganes, R., Gorrell, J.H., Gorrell, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M., Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kneitz, S., Kondziejewski, N., Kong, Y., Kovar, C., Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R., Marondel, I., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A., Stamps, A., Sugang, R., Tabor, P., Taylor, T., Vaequez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kucherlapati, R., Nelson, D. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (05-APR-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (23-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (28-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (27-DEC-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 191877)
Worley, K.C.
Direct Submission
Submitted (07-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 1999 this sequence version replaced gi:6091647.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlaj, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, J., Huiyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Sherty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, P., Umani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

Unpublished
Direct Submission
2 (bases 1 to 23162)
Rat Genome Sequencing Consortium.

REFERENCE

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KZWM

Center Clone name: CH230-unknown

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 133397 bases at least Q40

Consensus quality: 140476 bases at least Q30

Consensus quality: 145133 bases at least Q20

Estimated insert size: 122072; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 30 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 16165: contig of 16165 bp in length

* 16166 16265: gap of unknown length

* 16266 26917: contig of 10652 bp in length

* 26918 27017: gap of unknown length

* 27018 42777: contig of 15760 bp in length

* 42778 42877: gap of unknown length

* 42878 45839: contig of 2962 bp in length

* 45840 45939: gap of unknown length

* 45940 71248: contig of 25309 bp in length

* 71249 71348: gap of unknown length

* 71349 81456: contig of 10108 bp in length

* 81457 81556: gap of unknown length

* 81557 86328: contig of 4772 bp in length

* 86329 86428: gap of unknown length

* 86429 97963: contig of 11535 bp in length

* 97964 98063: gap of unknown length

* 98064 121383: contig of 23320 bp in length

* 121384 121483: gap of unknown length

* 121484 136774: contig of 15291 bp in length

* 136775 136874: gap of unknown length

* 136875 140310: contig of 3436 bp in length

* 140311 140410: gap of unknown length

* 140411 175013: contig of 34603 bp in length

* 175014 175113: gap of unknown length

* 175114 186168: contig of 11055 bp in length

* 186169 186269: gap of unknown length

* 186269 197528: contig of 11260 bp in length

* 197529 197628: gap of unknown length

* 197629 198730: contig of 1102 bp in length

* 198731 198830: gap of unknown length

* 198831 200482: contig of 1652 bp in length

* 200483 200582: gap of unknown length

* 200583 201947: contig of 1365 bp in length

* 201948 202047: gap of unknown length

* 202048 203310: contig of 1263 bp in length

* 203311 203410: gap of unknown length

* 203411 204930: contig of 1520 bp in length

* 204931 205030: gap of unknown length

* 205031 206032: contig of 1002 bp in length

* 206033 206132: gap of unknown length

* 206133 207315: contig of 1183 bp in length

* 207316 207415: gap of unknown length

* 207416 208596: contig of 1181 bp in length

* 208597 208696: gap of unknown length

* 208697 210100: contig of 1404 bp in length

* 210101 210200: gap of unknown length

* 210201 211444: contig of 1244 bp in length

* 211445 211544: gap of unknown length

* 211545 212545: contig of 1001 bp in length

* 212546 212645: gap of unknown length

* 212646 213804: contig of 1159 bp in length

* 213805 213804: contig of 1159 bp in length

```

* 213805 213904: gap of unknown length
* 213905 21523: contig of 1619 bp in length
* 21524 21523: gap of unknown length
* 21524 21523: gap of unknown length
* 217476 217476: contig of 1853 bp in length
* 217477 217576: gap of unknown length
* 217577 219822: contig of 2246 bp in length
* 219823 219822: gap of unknown length
* 219823 223162: contig of 3240 bp in length.

```

FEATURES

```

source
  1. .223162
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-unknown"
  1. .1156
    /note="wgs_contig"
    14247. .16165
    /note="wgs_contig"
    27018. .28348
    /note="wgs_contig"
    38658. .40255
    /note="wgs_contig"
    42878. .43969
    /note="wgs_contig"
    45940. .47659
    /note="wgs_contig"
    62225. .64598
    /note="wgs_contig"
    85049. .86328
    /note="wgs_contig"
    101184. .102685
    /note="wgs_contig"
    107615. .109278
    /note="wgs_contig"
    120385. .121383
    /note="wgs_contig"
    127313. .129027
    /note="wgs_contig"
    140411. .142152
    /note="wgs_contig"
    173831. .175013
    /note="wgs_contig"

```

```

Query Match      50.0%; Score 25; DB 2; Length 223162;
Best Local Similarity 69.4%; Pred. NO. 94;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

Qy 1 CCTACCCACCCCTGGATCCTTGGCCCGCCCTTGAATTCACACCTCCCA 49
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25824 CCTACACACCCACACATCTTGGCTGTCCCATCCCATCCCACTCCA 25872

```

```

Search completed: April 25, 2005, 12:18:04
Job time : 1097.17 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:06:08 ; Search time 498.707 Seconds
(without alignments)
593.509 Million cell updates/sec

Title: US-10-010-476-13
Perfect score: 50
Sequence: 1 CCTACCCACCTGATCTT.....TTGAATTCACACCTCCAC 50

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1990s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	50	2	AAX40728
2	30	60.0	30	2	AAX40725
3	30	60.0	30	2	AAX40724
4	26.6	53.2	1051	6	ABN74461 Bovine em
5	26.4	52.8	22444	9	ADA02591
6	26.4	52.8	22444	10	ADB72329
7	26.4	52.8	22444	10	ADB72329
8	25.2	50.4	422	4	AAK43752
9	25.2	50.4	422	6	ABT09033
10	25.2	50.4	705	12	ADG45731
11	25.2	50.4	4562	13	ADR15041
12	25.2	50.4	28564	10	ADD47107
13	25.2	50.4	28564	10	ADD47113
14	25.2	50.4	28564	10	ADD47117
15	25.2	50.4	28564	10	ADD47111
16	25.2	50.4	28564	10	ADD47111
17	25.2	50.4	28564	10	ADE83334
18	25.2	50.4	28564	10	ADE57550
19	25.2	50.4	28564	10	ADE57554
20	25	50.0	223556	11	ACN44110

94	22	44.0	2569	12	ADQ23193	Adq23193 Human sof	c 167	21.8	43.6	24757	4	AA333481	Aa333481 DNA encod
95	22	44.0	2774	13	ADQ39106	Adq39106 Human SNP	168	21.8	43.6	24757	10	ADB94490	Adb94490 Novel hum
96	22	44.0	2896	2	AA784537	Aa784537 Human kid	c 169	21.8	43.6	24875	10	ADD14677	Add14677 Human src
97	22	44.0	2896	10	AD331730	Ad331730 Human 946	c 170	21.8	43.6	34875	13	ADR52965	Adr52965 Drug ther
98	22	44.0	3392	13	ADRO7888	Adro7888 Full leng	c 171	21.8	43.6	34744	10	ADC19700	Adc19700 PMKADSHI
99	22	44.0	4164	8	AB276355	Ab276355 Human alp	c 172	21.8	43.6	54493	11	ACN45086	Acn45086 Human gen
100	22	44.0	4164	11	ADP65004	Adp65004 Human col	c 173	21.8	43.6	256493	11	ACN44514	Acn44514 Human gen
101	22	44.0	4164	11	ADP24472	Adp24472 PRO poly	c 174	21.8	43.6	257645	12	ADQ97289	Adq97289 Human can
102	22	44.0	4177	5	AA545077	Aa545077 cDNA enco	c 175	21.6	43.2	470	4	AAK66622	Aak66622 Human imm
103	22	44.0	4219	13	ACN40863	Acn40863 Tumour-as	c 176	21.6	43.2	478	4	AAK57028	Aak57028 Human imm
104	22	44.0	4603	8	ABX62997	Abx62997 Human act	177	21.6	43.2	526	4	AAK525328	Aak525328 Human ova
105	22	44.0	4690	12	ADQ25539	Adq25539 Human sof	178	21.6	43.2	540	12	ADL85957	Adl85957 DNA up-re
c 106	22	44.0	4839	5	ABV30270	Abv30270 Human pro	179	21.6	43.2	540	12	ADL85956	Adl85956 DNA up-re
c 107	22	44.0	5801	5	ABA18280	Abal8280 Human ner	c 180	21.6	43.2	564	6	ABQ27928	Abq27928 Oligonucl
c 108	22	44.0	8614	4	AAK74343	Aak74343 Human imm	181	21.6	43.2	564	6	ABQ27929	Abq27929 Oligonucl
c 109	22	44.0	8653	5	ABA15780	Abal5780 Human ner	c 182	21.6	43.2	592	13	ADQ50719	Adq50719 Novel can
c 110	22	44.0	46375	6	ABT10145	Abt10145 Human bre	c 183	21.6	43.2	705	3	AAZ53924	Aaz53924 Neibereria
c 111	22	44.0	46755	10	ADL13621	Adl13621 Osteoarth	c 184	21.6	43.2	716	12	ADJ74388	Adj74388 Rat cdna
c 112	22	44.0	97247	11	ACN44306	Acn44306 Human gen	c 185	21.6	43.2	732	10	ABZ39370	Abz39370 N. gonorr
c 113	22	44.0	110000	2	AAT42063	At42063 Continuation (2 of	186	21.6	43.2	828	4	AAK63505	Aak63505 Human imm
c 114	22	44.0	166910	12	ADN01278	Adn01278 Human end	187	21.6	43.2	1643	11	ADI31427	Adi31427 Human cdn
c 115	21.8	43.6	127	3	AA444698	Aa444698 Human sec	188	21.6	43.2	1680	3	AACT7967	Aac77967 Human can
c 116	21.8	43.6	296	3	AA293387	Aa293387 PCRScript	c 189	21.6	43.2	1824	6	AA516434	Aas16434 Human cdn
c 117	21.8	43.6	344	3	AACT74523	Aac74523 Human ORF	190	21.6	43.2	2475	4	AAK85125	Aak85125 Human imm
c 118	21.8	43.6	468	3	AA293386	Aa293386 PCRScript	c 191	21.6	43.2	2827	5	ADL45707	Adl45707 Human ova
c 119	21.8	43.6	531	4	AB2934502	Ab2934502 Human liv	192	21.6	43.2	3243	13	ACN39124	Acn39124 Tumour-as
c 120	21.8	43.6	566	3	AA293384	Aa293384 PCRScript	193	21.6	43.2	4041	4	ABK42635	Abk42635 Genomic s
c 121	21.8	43.6	570	4	ABA06910	Abao6910 Human pan	194	21.6	43.2	4041	9	ABE60791	Ab60791 Connectiv
c 122	21.8	43.6	570	4	AAK87892	Aak87892 Human dig	c 195	21.6	43.2	4221	4	ABK43171	Abk43171 DNA encod
c 123	21.8	43.6	573	3	AA293390	Aa293390 D14059 nu	c 196	21.6	43.2	4221	4	ABK43170	Abk43170 DNA encod
c 124	21.8	43.6	593	6	ABQ27948	Abq27948 Oligonucl	197	21.6	43.2	5575	3	AACT76044	Aac76044 Human ORF
c 125	21.8	43.6	593	6	ABQ27949	Abq27949 Oligonucl	198	21.6	43.2	6777	10	ADG42151	Adg42151 Human bra
c 126	21.8	43.6	665	6	ABQ25765	Abq25765 Oligonucl	c 199	21.6	43.2	13627	6	ABQ66975	Abq66975 Human ang
c 127	21.8	43.6	665	6	ABQ25764	Abq25764 Oligonucl	200	21.6	43.2	20001	13	ACN37218	Acn37218 Human per
c 128	21.8	43.6	699	3	AA297096	Aa297096 Human sec	201	21.6	43.2	24492	11	ACN44414	Acn44414 Human gen
c 129	21.8	43.6	699	9	ACH66725	Ach66725 Novel hum	202	21.6	43.2	37474	10	ADC19700	Adc19700 PMKADSHI
c 130	21.8	43.6	934	3	AA293389	Aa293389 PCRScript	c 203	21.6	43.2	40046	6	ABL68361	Ab168361 Kidney ca
c 131	21.8	43.6	1103	3	AA293388	Aa293388 PCRScript	c 204	21.6	43.2	40046	6	ABL68822	Ab168822 Kidney ca
c 132	21.8	43.6	1800	10	ADK64947	Aak64947 Disease t	c 205	21.6	43.2	40046	6	ABL68140	Ab168140 Kidney ca
c 133	21.8	43.6	1803	11	ACN91902	Acn91902 Breast ca	c 206	21.6	43.2	66681	13	ABD33333	Abd33333 Murine ca
c 134	21.8	43.6	2088	10	ADB62506	Adb62506 Human cdn	207	21.6	43.2	89378	12	ADN47591	Adn47591 20
c 135	21.8	43.6	2123	3	AA293391	Aa293391 D385848 n	208	21.6	43.2	89378	12	ADN47209	Adn47209 20
c 136	21.8	43.6	2170	8	ABT33964	Abt33964 Human pig	209	21.6	43.2	89378	12	ADN47960	Adn47960 20
c 137	21.8	43.6	2170	8	ABT33965	Abt33965 Human pig	c 210	21.6	43.2	110000	12	ADN46845	Adn46845 Thermoco
c 138	21.8	43.6	2170	8	ABT33959	Abt33959 Human pig	c 211	21.6	43.2	110000	12	ADN46123	Adn46123 Thermoco
c 139	21.8	43.6	2240	8	ABT33960	Abt33960 Human pig	c 212	21.6	43.2	110000	12	ADN46123	Adn46123 Thermoco
c 140	21.8	43.6	2240	10	ADC26792	Adc26792 Human lip	c 213	21.6	43.2	130244	13	ABD32872	Abd32872 Human can
c 141	21.8	43.6	2240	10	ADC26594	Adc26594 Human CYP	c 214	21.6	43.2	130244	13	ABD32872	Abd32872 Human can
c 142	21.8	43.6	2586	8	ACC46440	Acc46440 Human dit	c 215	21.6	43.2	131576	11	ACN44890	Acn44890 Human gen
c 143	21.8	43.6	2728	4	AAK94571	Aak94571 Human ful	216	21.6	43.2	203132	13	ABD33364	Abd33364 Murine ca
c 144	21.8	43.6	2728	12	ADL31453	Adl31453 Full leng	217	21.4	42.8	463	6	ABQ98389	Abq98389 Human ORF
c 145	21.8	43.6	3415	6	ABK34669	Abk34669 Human cdn	218	21.4	42.8	582	2	AAV86176	Aav86176 EST clone
c 146	21.8	43.6	4070	12	ADQ83444	Adq83444 Human tum	c 219	21.4	42.8	600	4	AAH70496	Aah70496 Human cer
c 147	21.8	43.6	4070	12	ADQ84677	Adq84677 Human tum	c 220	21.4	42.8	613	6	ABQ13834	Abq13834 Oligonucl
c 148	21.8	43.6	4070	13	ADQ85590	Adq85590 Human tum	221	21.4	42.8	613	6	ABQ13835	Abq13835 Oligonucl
c 149	21.8	43.6	4070	13	ADQ85590	Adq85590 Human tum	c 222	21.4	42.8	728	6	ABQ16160	Abq16160 Oligonucl
c 150	21.8	43.6	4500	8	AB225775	Ab225775 Human CYP	c 223	21.4	42.8	728	6	ABQ16161	Abq16161 Oligonucl
c 151	21.8	43.6	6306	4	AA454515	Aa454515 Chemicali	224	21.4	42.8	765	6	ABQ68823	Abq68823 Listeria
c 152	21.8	43.6	6306	6	ABK28457	Abk28457 DNA trans	225	21.4	42.8	783	4	AAH55733	Aah55733 Human bre
c 153	21.8	43.6	6472	6	ABQ72215	Abq72215 Human CYP	226	21.4	42.8	783	12	ADN40506	Adn40506 Human bre
c 154	21.8	43.6	6472	6	ABQ72364	Abq72364 Human CYP	227	21.4	42.8	990	2	AAZ19205	Aaz19205 M. tuberc
c 155	21.8	43.6	9432	6	AD34213	Ad34213 Human cyt	228	21.4	42.8	990	2	AAZ19417	Aaz19417 M. tuberc
c 156	21.8	43.6	9432	10	ACA61301	Ac61301 Human cyt	c 229	21.4	42.8	1058	3	AAO2099	Aao2099 Human col
c 157	21.8	43.6	9432	12	ADF83400	Adf83400 Human CYP	230	21.4	42.8	1097	2	AAV44325	Aav44325 Rat homol
c 158	21.8	43.6	9432	12	ADM78563	Adm78563 Human cyt	231	21.4	42.8	1097	10	ADBS2900	Adbs2900 Primary r
c 159	21.8	43.6	9432	12	ADM28891	Adm28891 Human wil	232	21.4	42.8	1101	4	AAH98767	Aah98767 Drosophil
c 160	21.8	43.6	9433	10	ACA61302	Ac61302 Human cyt	233	21.4	42.8	1101	10	ADE09534	Ade09534 Novel DNA
c 161	21.8	43.6	10211	2	AAV62152	Aav62152 HSV-2 str	234	21.4	42.8	1103	2	AAV42714	Aav42714 Rat homol
c 162	21.8	43.6	13677	12	ADJ78568	Adj78568 Human cyt	235	21.4	42.8	1268	10	ADE07277	Ade07277 Novel cod
c 163	21.8	43.6	13677	10	ADM28896	Adm28896 Human pse	236	21.4	42.8	1951	4	AAH43895	Aah43895 Human CLA
c 164	21.8	43.6	15766	12	ADE56804	Ade56804 Rat gen	237	21.4	42.8	1951	10	ADC22017	Adc22017 Human cad
c 165	21.8	43.6	18679	6	ABN80300	Abn80300 Human che	238	21.4	42.8	2378	4	AAH14851	Aah14851 Human cdn
c 166	21.8	43.6	24757	4	AA327687	Aa327687 DNA encod	239	21.4	42.8	3452	3	AAA37770	Aaa37770 Human TWI

240	21.4	42.8	3452	4	AA04872	Aad04872 Human TWI	C 313	21.2	42.4	987	6	ABQ50094	Abq50094 Oligonuc1
241	21.4	42.8	3514	3	AAA27747	Aaa27747 Human pot	C 314	21.2	42.4	997	4	AAH22162	Aah22162 Human mit
242	21.4	42.8	3514	13	ADR25089	Adr25089 Breast ca	C 315	21.2	42.4	1045	8	ABX63678	Abx63678 Human cDN
243	21.4	42.8	3514	13	ADR40138	Adr40138 Human TWI	C 316	21.2	42.4	1095	6	ABQ36090	Abq36090 Oligonuc1
244	21.4	42.8	3768	3	AAA15953	Aaa15953 Human pro	C 317	21.2	42.4	1095	6	ABQ36091	Abq36091 Oligonuc1
245	21.4	42.8	3912	13	ABD32908	Abd32908 Human can	C 318	21.2	42.4	1164	6	ABQ17873	Abq17873 Oligonuc1
246	21.4	42.8	4708	4	AAK73842	Aak73842 Human imm	C 319	21.2	42.4	1164	6	ABQ17872	Abq17872 Oligonuc1
247	21.4	42.8	6172	6	ABL34572	Ab134572 Human met	C 320	21.2	42.4	1235	4	AAK73842	Aak73842 Human can
248	21.4	42.8	6172	7	ADS99833	Ades99833 Bisulphit	C 321	21.2	42.4	1235	4	AAK73842	Aak73842 Human can
249	21.4	42.8	7225	10	ADSR5178	Ades5178 Farnesyl	C 322	21.2	42.4	1235	10	ADE47278	Ades47278 Human car
250	21.4	42.8	7225	10	ADSR5178	Ades5178 Farnesyl	C 323	21.2	42.4	1235	10	ADE47278	Ades47278 Human car
251	21.4	42.8	10024	6	ABN95933	Abn95933 Gene #243	C 324	21.2	42.4	1235	13	ADJ08696	Adj08696 Human car
252	21.4	42.8	10373	11	ACN44293	Acn44293 Mouse mRN	C 325	21.2	42.4	1235	13	ADJ08695	Adj08695 Human car
253	21.4	42.8	10373	11	ACN44293	Acn44293 Mouse mRN	C 326	21.2	42.4	1294	4	AAI60661	Aai60661 Human pol
254	21.4	42.8	60915	13	ABD33520	Abd33520 Human can	C 327	21.2	42.4	1294	4	AAI60661	Aai60661 Human pol
255	21.4	42.8	60915	13	ABD32907	Abd32907 Human can	C 328	21.2	42.4	1294	10	ADC32388	Adc32388 Human nov
256	21.4	42.8	81005	13	ABD33213	Abd33213 Murine ca	C 329	21.2	42.4	1316	4	AAI58875	Aai58875 Human pol
257	21.4	42.8	93483	9	ADA03083	Ada03083 Mouse mCG	C 330	21.2	42.4	1316	5	ADQ99097	Adq99097 DNA encod
258	21.4	42.8	93483	9	ADA66367	Ada66367 Mouse mCG	C 331	21.2	42.4	1316	9	AD848857	Ad848857 Novel hum
259	21.4	42.8	93483	10	ADB72821	Adb72821 Mouse mCG	C 332	21.2	42.4	1322	4	AAK727733	Aak727733 Human tra
260	21.4	42.8	93483	11	ADL27161	Adl27161 Mouse gen	C 333	21.2	42.4	1336	3	AAK727733	Aak727733 Human tra
261	21.4	42.8	100660	13	ABD32721_3	Abd32721_3 Continuation (4 of	C 334	21.2	42.4	1351	4	AAH19225	Aah19225 Human sec
262	21.4	42.8	102374	11	ACN44292	Acn44292 Mouse gen	C 335	21.2	42.4	1361	13	ADT46303	Adt46303 Bacterial
263	21.4	42.8	110000	11	ACN43984_2	Acn43984_2 Continuation (3 of	C 336	21.2	42.4	1418	3	AAK74843	Aak74843 Human ORF
264	21.4	42.8	304905	11	ADP75180_2	Adp75180_2 Human end	C 337	21.2	42.4	1563	12	ADQ95987	Adq95987 T cell ac
265	21.4	42.8	310268	13	ABD32548	Abd32548 Human can	C 338	21.2	42.4	1602	4	ADQ5248	Adq5248 Human sec
266	21.2	42.4	311	6	ABN75105	Abn75105 Human ORF	C 339	21.2	42.4	1633	10	ABQ62892	Abq62892 Human cDN
267	21.2	42.4	324	6	ABN79462	Abn79462 Human ORF	C 340	21.2	42.4	1747	6	ABQ18264	Abq18264 Oligonuc1
268	21.2	42.4	351	4	AAK35544	Aak35544 Human car	C 341	21.2	42.4	1747	10	ABQ18265	Abq18265 Oligonuc1
269	21.2	42.4	351	10	ADL45623	Adl45623 Human car	C 342	21.2	42.4	1852	10	ABS58306	Abs58306 Spider dr
270	21.2	42.4	351	13	ADJ07041	Adj07041 Human car	C 343	21.2	42.4	2004	2	AAK785356	Aak785356 Nephila c
271	21.2	42.4	356	12	ADQ77047	Adq77047 Partner d	C 344	21.2	42.4	2034	11	ABD13354	Abd13354 Pseudomon
272	21.2	42.4	357	12	ADL92440	Adl92440 Partner d	C 345	21.2	42.4	2037	6	AAK62787	Aak62787 cDNA sequ
273	21.2	42.4	357	12	ADL92450	Adl92450 Partner d	C 346	21.2	42.4	2060	11	ADM02387	Adm02387 Human cDN
274	21.2	42.4	357	12	ADL92456	Adl92456 Partner d	C 347	21.2	42.4	2082	8	ABX71139	Abx71139 Novel hum
275	21.2	42.4	357	12	ADQ77057	Adq77057 Partner d	C 348	21.2	42.4	2089	13	ADS48370	Ads48370 Bacterial
276	21.2	42.4	357	12	ADQ77061	Adq77061 TAR2h-5d1	C 349	21.2	42.4	2203	3	AAK33362	Aak33362 Human sec
277	21.2	42.4	357	12	ADQ77063	Adq77063 TAR2h-5d1	C 350	21.2	42.4	2234	13	ADQ86958	Adq86958 Human tum
278	21.2	42.4	360	12	ADL92396	Adl92396 Single do	C 351	21.2	42.4	2268	5	AAK85432	Aak85432 DNA encod
279	21.2	42.4	360	12	ADQ77002	Adq77002 TAR2-10 c	C 352	21.2	42.4	2268	6	ABS64952	Abs64952 Human cDN
280	21.2	42.4	368	12	ADQ77069	Adq77069 TAR2h-10 c	C 353	21.2	42.4	2276	4	AAK21248	Aak21248 Human cDN
281	21.2	42.4	381	13	ADK78338	Adk78338 TAR2-10-2	C 354	21.2	42.4	2276	8	ACA03607	ACA03607 CDNA enco
282	21.2	42.4	418	13	ACN52675	Acn52675 Cotton an	C 355	21.2	42.4	2276	8	AXH89145	Axh89145 DNA encod
283	21.2	42.4	443	13	ACN55788	Acn55788 Cotton an	C 356	21.2	42.4	2276	8	ACD41799	ACD41799 Human sec
284	21.2	42.4	448	9	ACH31810	Ach31810 Human end	C 357	21.2	42.4	2276	8	ACA04028	ACA04028 Human cDN
285	21.2	42.4	456	9	ACH36695	Ach36695 Human end	C 358	21.2	42.4	2276	9	ADA75959	Ada75959 Human PRO
286	21.2	42.4	465	9	ACH21869	Ach21869 Human adu	C 359	21.2	42.4	2276	9	ADA18609	Ada18609 Human PRO
287	21.2	42.4	467	4	AAK68466	Aak68466 Human imm	C 360	21.2	42.4	2276	9	ADA61232	Ada61232 Homo sapi
288	21.2	42.4	470	9	ACH32324	Ach32324 Human imm	C 361	21.2	42.4	2276	9	ADB19017	Adb19017 Novel hum
289	21.2	42.4	473	9	ACH36008	Ach36008 Human end	C 362	21.2	42.4	2276	9	ADB27558	Adb27558 CDNA enco
290	21.2	42.4	477	11	ABD13315	Abd13315 Pseudomon	C 363	21.2	42.4	2276	9	ADA86037	Ada86037 Novel hum
291	21.2	42.4	505	6	ABQ48305	Abq48305 Oligonuc1	C 364	21.2	42.4	2276	9	ADB15601	Adb15601 Human PRO
292	21.2	42.4	505	6	ABQ48304	Abq48304 Oligonuc1	C 365	21.2	42.4	2276	9	ADA73787	Ada73787 Human PRO
293	21.2	42.4	526	6	ABQ36055	Abq36055 Oligonuc1	C 366	21.2	42.4	2276	9	ADA67182	Ada67182 Human PRO
294	21.2	42.4	526	6	ABQ36054	Abq36054 Oligonuc1	C 367	21.2	42.4	2276	9	ADA67182	Ada67182 Human PRO
295	21.2	42.4	546	6	ABQ26184	Abq26184 Oligonuc1	C 368	21.2	42.4	2276	9	ADA85485	Ada85485 Novel hum
296	21.2	42.4	546	6	ABQ26185	Abq26185 Oligonuc1	C 369	21.2	42.4	2276	9	ADA96687	Ada96687 Human PRO
297	21.2	42.4	549	4	AAK24885	Aak24885 Human ova	C 370	21.2	42.4	2276	9	ADA79001	Ada79001 Human PRO
298	21.2	42.4	599	5	AAH83525	Aah83525 Human ova	C 371	21.2	42.4	2276	9	ADA87140	Ada87140 Novel hum
299	21.2	42.4	629	13	ACN54594	Acn54594 Cotton an	C 372	21.2	42.4	2276	9	ADB16342	Adb16342 Human PRO
300	21.2	42.4	648	12	ADK52083	Adk52083 Human ato	C 373	21.2	42.4	2276	9	ADA91434	Ada91434 Novel hum
301	21.2	42.4	653	6	ABQ41380	Abq41380 Oligonuc1	C 374	21.2	42.4	2276	9	ADB14497	Adb14497 Human PRO
302	21.2	42.4	653	6	ABQ41381	Abq41381 Oligonuc1	C 375	21.2	42.4	2276	9	ADB18458	Adb18458 Novel hum
303	21.2	42.4	703	8	AAK55716	Aak55716 Nephila m	C 376	21.2	42.4	2276	9	ADA93673	Ada93673 Human PRO
304	21.2	42.4	707	6	ABQ16538	Abq16538 Oligonuc1	C 377	21.2	42.4	2276	9	ADB19569	Adb19569 Novel hum
305	21.2	42.4	707	6	ABQ16539	Abq16539 Oligonuc1	C 378	21.2	42.4	2276	9	ADB12881	Adb12881 Human PRO
306	21.2	42.4	710	2	AAZ15876	Aaz15876 Human gen	C 379	21.2	42.4	2276	9	ACD98428	ACD98428 Novel hum
307	21.2	42.4	740	5	ADM19417	Adm19417 Novel hum	C 380	21.2	42.4	2276	9	ADB24135	Adb24135 Human PRO
308	21.2	42.4	928	6	ABQ34275	Abq34275 Oligonuc1	C 381	21.2	42.4	2276	9	ADA74368	Ada74368 Human PRO
309	21.2	42.4	928	6	ABQ34274	Abq34274 Oligonuc1	C 382	21.2	42.4	2276	9	ADA81892	Ada81892 Human PRO
310	21.2	42.4	941	6	ABQ51076	Abq51076 Oligonuc1	C 383	21.2	42.4	2276	9	ADA74855	Ada74855 Human PRO
311	21.2	42.4	941	6	ABQ51077	Abq51077 Oligonuc1	C 384	21.2	42.4	2276	9	ADA84933	Ada84933 Novel hum
312	21.2	42.4	987	6	ABQ50095	Abq50095 Oligonuc1	C 385	21.2	42.4	2276	9	ADB29637	Adb29637 cDNA enco

C 386	21.2	42.4	2276	9	ADA80165	AdA80165	Human PRO
C 387	21.2	42.4	2276	9	ADA75407	Human PRO	
C 388	21.2	42.4	2276	9	ADA46632	Human PRO	
C 389	21.2	42.4	2276	9	ADB24928	Human PRO	
C 390	21.2	42.4	2276	9	ADA93104	Human PRO	
C 391	21.2	42.4	2276	9	ADB26454	cDNA enco	
C 392	21.2	42.4	2276	9	ADB30741	cDNA enco	
C 393	21.2	42.4	2276	9	ADA60669	Homo sapi	
C 394	21.2	42.4	2276	9	ADB23816	Human PRO	
C 395	21.2	42.4	2276	9	ADA96145	Human PRO	
C 396	21.2	42.4	2276	9	ADA80717	Human PRO	
C 397	21.2	42.4	2276	9	ADA95593	Human PRO	
C 398	21.2	42.4	2276	9	ADB25902	cDNA enco	
C 399	21.2	42.4	2276	9	ADB21387	Novel hum	
C 400	21.2	42.4	2276	9	ADA77166	Human PRO	
C 401	21.2	42.4	2276	9	ADB17906	cDNA enco	
C 402	21.2	42.4	2276	9	ADA86589	Novel hum	
C 403	21.2	42.4	2276	9	ADA87692	Novel hum	
C 404	21.2	42.4	2276	9	ADA46080	Novel hum	
C 405	21.2	42.4	2276	9	ADB28110	cDNA enco	
C 406	21.2	42.4	2276	9	ADB28662	cDNA enco	
C 407	21.2	42.4	2276	9	ADA76614	Human PRO	
C 408	21.2	42.4	2276	9	ADA88244	Novel hum	
C 409	21.2	42.4	2276	9	ADA97249	Human PRO	
C 410	21.2	42.4	2276	9	ADB27006	cDNA enco	
C 411	21.2	42.4	2276	9	ADB21939	Novel hum	
C 412	21.2	42.4	2276	9	ADA66630	Human PRO	
C 413	21.2	42.4	2276	9	ADB22491	Human PRO	
C 414	21.2	42.4	2276	9	ADB23264	Human PRO	
C 415	21.2	42.4	2276	9	ADA91986	Novel hum	
C 416	21.2	42.4	2276	9	ADB15049	Human PRO	
C 417	21.2	42.4	2276	9	ADB38301	Novel hum	
C 418	21.2	42.4	2276	9	ADB37749	Novel hum	
C 419	21.2	42.4	2276	9	ADB66221	Novel hum	
C 420	21.2	42.4	2276	10	ADB89301	Human PRO	
C 421	21.2	42.4	2276	10	ADB90033	Human PRO	
C 422	21.2	42.4	2276	10	ADB39134	Novel hum	
C 423	21.2	42.4	2276	10	ADB46757	Novel hum	
C 424	21.2	42.4	2276	10	ADB86364	Human PRO	
C 425	21.2	42.4	2276	10	ADB76969	Novel hum	
C 426	21.2	42.4	2276	10	ADB34126	Human PRO	
C 427	21.2	42.4	2276	10	ADB35230	Human PRO	
C 428	21.2	42.4	2276	10	ADB33574	Human PRO	
C 429	21.2	42.4	2276	10	ADB34678	Human PRO	
C 430	21.2	42.4	2276	10	ADB35782	Human PRO	
C 431	21.2	42.4	2276	10	ADB46177	Novel hum	
C 432	21.2	42.4	2276	10	ADC50050	Novel hum	
C 433	21.2	42.4	2276	10	ADC71597	Novel hum	
C 434	21.2	42.4	2276	10	ADC59576	Novel hum	
C 435	21.2	42.4	2276	10	ADC52583	Novel hum	
C 436	21.2	42.4	2276	10	ADC56937	Novel hum	
C 437	21.2	42.4	2276	10	ADC60128	Novel hum	
C 438	21.2	42.4	2276	10	ADC50603	Novel hum	
C 439	21.2	42.4	2276	10	ADC65130	Human PRO	
C 440	21.2	42.4	2276	10	ADC54228	Novel hum	
C 441	21.2	42.4	2276	10	ADC53189	Novel hum	
C 442	21.2	42.4	2276	10	ADC58712	Novel hum	
C 443	21.2	42.4	2276	10	ADC55590	Novel hum	
C 444	21.2	42.4	2276	10	ADC58160	Novel hum	
C 445	21.2	42.4	2276	10	ADD02834	Novel hum	
C 446	21.2	42.4	2276	10	ADC89826	Novel hum	
C 447	21.2	42.4	2276	10	ADC69245	cDNA enco	
C 448	21.2	42.4	2276	10	ADC48134	Human PRO	
C 449	21.2	42.4	2276	10	ADD09663	Human PRO	
C 450	21.2	42.4	2276	10	ADD04238	Novel hum	
C 451	21.2	42.4	2276	10	ADC80194	Novel hum	
C 452	21.2	42.4	2276	10	ADD10701	Human PRO	
C 453	21.2	42.4	2276	10	ADC47582	Human PRO	
C 454	21.2	42.4	2276	10	ADC79642	Novel hum	
C 455	21.2	42.4	2276	10	ADD09111	Human PRO	
C 456	21.2	42.4	2276	10	ADD40824	Novel hum	
C 457	21.2	42.4	2276	10	ADD51963	cDNA enco	
C 458	21.2	42.4	2276	10	ADD52703	cDNA enco	

C 459	21.2	42.4	2276	10	ADD53255	Novel hum	
C 460	21.2	42.4	2276	10	ADD51411	cDNA enco	
C 461	21.2	42.4	2276	10	ADD02210	Human PRO	
C 462	21.2	42.4	2276	10	ADD01644	Human PRO	
C 463	21.2	42.4	2276	10	ADD53826	Novel hum	
C 464	21.2	42.4	2276	10	ADD92143	Human PRO	
C 465	21.2	42.4	2276	10	ADD91039	Human PRO	
C 466	21.2	42.4	2276	10	ADD03653	Human PRO	
C 467	21.2	42.4	2276	10	ADD31950	Novel hum	
C 468	21.2	42.4	2276	10	ADD21882	cDNA enco	
C 469	21.2	42.4	2276	10	ADD79106	cDNA enco	
C 470	21.2	42.4	2276	10	ADD41642	Human PRO	
C 471	21.2	42.4	2276	10	ADD17459	Human PRO	
C 472	21.2	42.4	2276	10	ADD91591	Human PRO	
C 473	21.2	42.4	2276	10	ADD91591	Human PRO	
C 474	21.2	42.4	2276	10	ADD33054	Novel hum	
C 475	21.2	42.4	2276	10	ADD33606	Novel hum	
C 476	21.2	42.4	2276	10	ADD79658	cDNA enco	
C 477	21.2	42.4	2276	10	ADD92695	Human PRO	
C 478	21.2	42.4	2276	10	ADD19115	Human PRO	
C 479	21.2	42.4	2276	10	ADD18563	Human PRO	
C 480	21.2	42.4	2276	10	ADD42759	Human PRO	
C 481	21.2	42.4	2276	10	ADD95548	Human PRO	
C 482	21.2	42.4	2276	10	ADD22434	cDNA enco	
C 483	21.2	42.4	2276	10	ADD78552	cDNA enco	
C 484	21.2	42.4	2276	10	ADD42194	Human PRO	
C 485	21.2	42.4	2276	10	ADD80210	cDNA enco	
C 486	21.2	42.4	2276	10	ADD89238	Human PRO	
C 487	21.2	42.4	2276	10	ADD40522	Human PRO	
C 488	21.2	42.4	2276	10	ADD04321	Human PRO	
C 489	21.2	42.4	2276	10	ADD92450	Human PRO	
C 490	21.2	42.4	2276	10	ADG21159	Novel hum	
C 491	21.2	42.4	2276	10	ADG32800	Novel hum	
C 492	21.2	42.4	2276	10	ADF97135	Human PRO	
C 493	21.2	42.4	2276	10	ADG80199	Human PRO	
C 494	21.2	42.4	2276	10	ADG79647	Human PRO	
C 495	21.2	42.4	2276	10	ADH54939	Novel hum	
C 496	21.2	42.4	2276	10	ADH55491	Novel hum	
C 497	21.2	42.4	2276	10	ADI64659	Novel hum	
C 498	21.2	42.4	2276	10	ADI63158	Novel hum	
C 499	21.2	42.4	2276	10	ADH81572	Novel hum	
C 500	21.2	42.4	2276	10	ADH81020	Novel hum	

ALIGNMENTS

RESULT 1

AAAX40728
ID AAX40728 standard; DNA; 50 BP.

AC AAX40728;

XX

DT 16-JUN-1999 (first entry)

XX GC-box c-MET nucleotide sequence.

DE DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;

KW allosteric site; synthetic inhibitor; cancer cell; proliferation;

KW development disorder; Huntington's disease; Down's syndrome; Hox gene;

XX ss.

XX Synthetic.

OS WO9912027-A1.

XX 11-MAR-1999.

XX 12-JUN-1998; 98WO-US012351.

XX 29-AUG-1997; 97US-0057411P.

XX (REGC) UNIV CALIFORNIA.

PA

XX PI Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Claim 10; Page 63; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase), thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 50 BP; 8 A; 27 C; 5 G; 10 T; 0 U; 0 Other;
 XX
 XX Query Match 100.0%; Score 50; DB 2; Length 50;
 XX Best Local Similarity 100.0%; Pred. No. 5.6e-08;
 XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 1 CCTACCCACCCTGGATCCTTGGCCGCCGCCCTTGAATTCCTCCAC 50
 XX 1 CCTACCCACCCTGGATCCTTGGCCGCCGCCCTTGAATTCCTCCAC 50
 XX
 XX RESULT 2
 XX AAX40725
 XX ID AAX40725 standard; DNA; 30 BP.
 XX AC AAX40725;
 XX DT 16-JUN-1999 (first entry)
 XX GC-box b-MET nucleotide sequence.
 XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 XX Synthetic.
 XX WO9912027-A1.
 XX 11-MAR-1999.
 XX 16-JUN-1999 (first entry)
 XX GC-box b-MET nucleotide sequence.
 XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 XX Synthetic.
 XX WO9912027-A1.
 XX 11-MAR-1999.
 XX 12-JUN-1998; 98WO-US012351.
 XX 29-AUG-1997; 97US-0057411P.
 XX (REGC) UNIV CALIFORNIA.
 XX Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Claim 10; Page 39; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase), thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification

CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 30 BP; 3 A; 14 C; 5 G; 8 T; 0 U; 0 Other;
 XX
 XX Query Match 60.0%; Score 30; DB 2; Length 30;
 XX Best Local Similarity 100.0%; Pred. No. 0.51;
 XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 11 CTGGATCCTTGGCCGCCGCCCTTGAATTCCTCC 40
 XX 1 CTGGATCCTTGGCCGCCGCCCTTGAATTCCTCC 30
 XX
 XX RESULT 3
 XX AAX40724/c
 XX ID AAX40724 standard; DNA; 30 BP.
 XX AC AAX40724;
 XX DT 16-JUN-1999 (first entry)
 XX GC-box a nucleotide sequence of the specification.
 XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 XX Synthetic.
 XX WO9912027-A1.
 XX 11-MAR-1999.
 XX 12-JUN-1998; 98WO-US012351.
 XX 29-AUG-1997; 97US-0057411P.
 XX (REGC) UNIV CALIFORNIA.
 XX Reich NO, Flynn J;
 XX WPI; 1999-205256/17.
 XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX Example 2; Page 39; 114pp; English.
 XX The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DNMTase), thereby modulating DNMTase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX Sequence 30 BP; 8 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
 XX
 XX Query Match 60.0%; Score 30; DB 2; Length 30;
 XX Best Local Similarity 100.0%; Pred. No. 0.51;
 XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 11 CTGGATCCTTGGCCGCCGCCCTTGAATTCCTCC 40
 XX 30 CTGGATCCTTGGCCGCCGCCCTTGAATTCCTCC 1

Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast; prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; gene; ds.

KW

RESULT 4
ABN74461/c

KW

ID ABN74461 standard; cDNA; 1051 BP.

KW

XX AC ABN74461;

XX

DT 03-JUL-2002 (first entry)

XX

DE Bovine embryonic germ (EG) cell cDNA EST #512.

XX

KW Bovine; Bos taurus; EST; expressed sequence tag; totipotence;

XX

KW development; gene; ss.

XX

XX Bos taurus.

XX

OS WO200194550-A2.

PN

XX 13-DEC-2001.

XX

PF 07-JUN-2001; 2001WO-US018576.

XX

PR 07-JUN-2000; 2000US-0209874P.

XX

PR 06-JUN-2001; 2001US-00876143.

XX

XX (INFI-) INFIGEN INC.

PA

PI Eilertsen KJ, Pfister-Genskow M, Childs L;

PI

XX WPI; 2002-351289/38.

XX

XX An expressed sequence tag (EST), the expression of which, or its

XX

PT complementary sequence, in a cell identifies the cell as a

PT

PT developmentally competent or incompetent cell.

PT

XX Example 16; Page 441-442; 584pp; English.

XX

XX The present invention describes an expressed sequence tag (EST), where the EST is an isolated, enriched, or purified nucleic acid sequence representing all or part of a gene, the expression of which, or its complementary sequence, in a cell identifies the cell as a developmentally competent or incompetent cell. Molecules which induce developmental competence in a cell line are useful for inducing totipotence in one or more cells. Molecules which induce developmental competence in a cell line are useful for preventing a full term pregnancy in an animal and inhibiting totipotence. The molecules are also useful for treating a disease in an animal by inducing development of one or more cells of the animal into a specific cell type. The present sequence represents a bovine EST which is given in the exemplification of the present invention

XX

SQ Sequence 1051 BP; 208 A; 279 C; 242 G; 236 T; 0 U; 86 Other;

SQ

Query Match 53.2%; Score 26.6; DB 6; Length 1051;

Query Match

Best Local Similarity 66.7%; Pred. No. 12;

Best Local Similarity

Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Matches

QY 1 CCTACCCACCCCTGGATCTTGGCCGCCCTTGAATTCACACCTCC 48

QY

DB 767 CCCTCCNATNAANNTTCTTGGCCGCCCTTGAATTCACACCTCC 720

DB

RESULT 5

ADA02591

RESULT 6

ID ADA02591 standard; DNA; 22444 BP.

ID

XX ADA02591;

XX

DT 06-NOV-2003 (first entry)

XX

XX Mouse Bml1 carcinoma associated gene, SEQ ID NO:1109.

XX

DE

OS

Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast; prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; gene; ds.

OS

XX Mus sp.

XX

PN WO2003057146-A2.

PN

XX 17-JUL-2003.

XX

PF 26-DEC-2002; 2002WO-US041414.

XX

PR 26-DEC-2001; 2001US-00035832.

XX

XX (SAGR-) SAGRES DISCOVERY.

XX

XX Morris DW;

XX

XX WPI; 2003-587068/55.

XX

XX New recombinant nucleic acid encoding carcinoma associated protein,

XX

XX useful for preparing compositions for treating carcinomas.

XX

PS Claim 1; SEQ ID NO 1109; 245pp; English.

PS

XX The invention relates to recombinant carcinoma associated (CA) nucleic acid sequences from mouse and human (ADA01482-ADA03094), and to recombinant carcinoma associated proteins (CAP) encoded by them. The invention also encompasses expression vectors and host cells comprising a CA nucleic acid, a polypeptide (especially an antibody) that specifically binds to the protein, and a biochip comprising CA nucleic acid or fragments thereof. The sequences of the invention were identified using oncogenic retroviruses, which insert into the genome of the host organism at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose carcinoma (especially breast cancer, prostate cancer, lymphoma or leukaemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed murine CA nucleic acid sequence of the invention. Note: The complete sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 22444 BP; 4418 A; 3243 C; 3397 G; 4846 T; 0 U; 6540 Other;

SQ

Query Match 52.8%; Score 26.4; DB 9; Length 22444;

Query Match

Best Local Similarity 75.0%; Pred. No. 20;

Best Local Similarity

Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Matches

QY 7 CACCTGGATCTTGGCCGCCCTTGAATTCACACCTCCAC 50

QY

DB 9144 CACCTGGTCTTGGCCGCCCTTGAATTCACACCTCCAC 9187

DB

RESULT 6

RESULT 6

ID ADB72329 standard; DNA; 22444 BP.

ID

XX ADB72329;

XX

XX 04-DEC-2003 (first entry)

XX

XX Mouse Bml1 gene.

XX

XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

XX

XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX

XX

XX

XX

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 422 BP; 158 A; 57 C; 124 G; 83 T; 0 U; 0 Other;

Query Match 50.4%; Score 25.2; DB 4; Length 422;
Best Local Similarity 78.9%; Pred. No. 33;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGATCTTGGCCCGCCCTTGAATTCCTCCCAACCTC 47
DB 159 CCTGTCTCTGACCCCTGTCTTGTGATTCCTCCCAACCTC 122

RESULT 9
AAK30749/c
ID AAK30749 standard; DNA; 422 BP.
XX
AC AAK30749;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 5306.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-489900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 5306; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 422 BP; 158 A; 57 C; 124 G; 83 T; 0 U; 0 Other;

Query Match 50.4%; Score 25.2; DB 4; Length 422;
Best Local Similarity 78.9%; Pred. No. 33;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGATCTTGGCCCGCCCTTGAATTCCTCCCAACCTC 47
DB 159 CCTGTCTCTGACCCCTGTCTTGTGATTCCTCCCAACCTC 122

RESULT 10

ABT09033
ID ABT09033 standard; DNA; 705 BP.
XX
AC ABT09033;
XX
DT 05-DEC-2002 (first entry)
XX
DE Phase-1 Rat CT gene SEQ ID No 121.
XX
KW Rat; toxicity study; rat toxic response gene; toxicological response;
KW drug development; phase-1 rat CT gene; ds.
XX
OS Rattus sp.
XX
PN WO200266682-A2.
XX
PD 29-AUG-2002.
XX
PF 29-JAN-2002; 2002WO-US002935.
XX
PR 29-JAN-2001; 2001US-0264933P.
PR 26-JUL-2001; 2001US-0308161P.
XX
PA (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX
PI Farris G, Hicken SH, Farr SB;
XX
DR WPI; 2002-674961/72.
XX
PT Evaluating the toxicity of an agent, useful in drug development or in
PT determining toxicological responses to a new drug, by determining the
PT expression of rat toxicologically relevant genes in the test animal in
PT response to the test agent.
XX
PS Disclosure; Page 135; 388pp; English.
XX
CC The invention relates to a method used for evaluating the toxicity of an
CC agent comprising determining the expression of a rat toxic response
CC gene(s) in the test animal in response to the agent. The method is useful
CC in drug development, particularly for conducting toxicity studies and
CC analysis before a new drug or compound is approved for human consumption
CC or use. The method is also useful in determining toxicological responses
CC to a new drug. This polynucleotide sequence represents a phase-1 rat CT
CC gene of the invention
XX
SQ Sequence 705 BP; 180 A; 187 C; 181 G; 155 T; 0 U; 2 Other;

Query Match 50.4%; Score 25.2; DB 6; Length 705;
Best Local Similarity 78.9%; Pred. No. 35;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCTCCCGCCCTTGAATTC 38
DB 175 CCTATCGTGTGGATCCTTGCCTCCCGCCCTGAGTTC 212

RESULT 11
ADG45731
ID ADG45731 standard; DNA; 705 BP.
XX
AC ADG45731;
XX
DT 26-FEB-2004 (first entry)
XX
DE Liver inflammatory predictive gene related DNA sequence.
XX
KW liver; liver toxicity; liver toxicity predictive gene;
KW liver inflammation predictive gene; inflammation; gene; ds; rat.
XX
OS Rattus norvegicus.
XX
PN WO2003095624-A2.
XX

```

PD 20-NOV-2003.
XX
PF 09-MAY-2003; 2003MO-US014832.
XX
PR 10-MAY-2002; 2002US-0379831P.
XX
PA (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX
PI Kier L, Nolan TD, Sankar U, Derbel M;
XX
DR WPI; 2004-042456/04.
XX
PT Predicting the liver toxicity in an individual to an agent by measuring
PT the expression of one or more liver toxicity predictive genes in the
PT sample from the individual treated with the agent.
XX
PS Claim 7; Fig 4; 336pp; English.
XX
CC The present invention describes a method for predicting the liver
CC toxicity in an individual to an agent. The method comprises: (a)
CC obtaining a biological sample from the individual treated with the agent;
CC (b) measuring the expression of one or more liver toxicity predictive
CC genes in the sample; and (c) using the test expression profile with a set
CC of reference expression profiles in a Predictive Model to determine
CC whether the agent will induce liver toxicity in the individual. Also
CC described: (1) predicting the liver toxicity of an agent using an in
CC vitro system; (2) predicting the liver toxicity in a biological sample
CC from an individual, in vitro cell cultures or explants to an agent via a
CC programmable machine; (3) a computer program product for enabling a
CC computer to perform Predictive Model analysis for liver toxicity on a
CC biological sample from an individual, in vitro cell cultures or explants
CC to an agent; (4) a computer system adopted to predict liver toxicity in a
CC biological system from an individual, in vitro cell cultures or explants
CC to an agent; (5) a computer program product for predicting liver toxicity
CC from a test sample expression profile; (6) mining genes predictive of
CC liver toxicity; (7) an integrated system for predicting liver toxicity;
CC and (8) identifying one or more liver inflammation predictive genes. The
CC method is useful for predicting the liver toxicity in an individual to an
CC agent. The present sequence represents a liver toxicity predictive gene
CC sequence that represents a 24 hour combo 3 gene, which is specifically
CC claimed in the exemplification of the present invention.
XX
SQ Sequence 705 BP; 180 A; 187 C; 181 G; 155 T; 0 U; 2 Other;

Query Match 50.4%; Score 25.2; DB 12; Length 705;
Best Local Similarity 78.9%; Pred. No. 35;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTACCCACCCTGGATCTTGGCCCGCCCTTGAATTC 38
DB 175 CCTATCGTGGCTGGATCTTGGCCCGCCACACCTTGAGTTC 212

RESULT 12
ADRI5041
ID ADRI5041 standard; DNA; 4562 BP.
XX
AC ADRI5041;
XX
XX 21-OCT-2004 (first entry)
XX
DE Rat electroconvulsive seizure (EC) signature gene seqid 74.
XX
KW neuroleptic; nootropic; antidepressant; eating disorder; tranquilizer;
KW neuropsychiatric disorder; signature gene; electroconvulsive seizure;
KW ECS; schizophrenia; autism; major depressive disorder; MDD;
KW bipolar affective disorder; BAD; psychotic depression; anxiety disorder;
KW eating disorder; attention deficit hyperactivity disorder;
KW neuropsychiatric drug; gene; ss; rat.
XX
OS Rattus norvegicus.
XX
PN US2004152107-A1.

```

```

XX
PD 05-AUG-2004.
XX
PF 18-SEP-2003; 2003US-00664705.
XX
PR 18-SEP-2002; 2002US-0411718P.
PR 18-SEP-2002; 2002US-0431882P.
PR 18-SEP-2002; 2002US-0431882P.
PR 18-JUN-2003; 2003US-0479970P.
XX
PA (ALTA/) ALTAR C A.
PA (LAEN/) LAENG P.
PA (YOUN/) YOUNG T A.
PA (CHAR/) CHARLES V D.
XX
PI Altar CA, Laeng P, Young TA, Charles VD;
XX
DR WPI; 2004-580183/56.
XX
PT Identifying compound to treat neuropsychiatric disorder, by contacting
PT cell with test compound, determining expression of signature genes by
PT cell, comparing expression of signature genes to expression in cell not
PT contacted with compound.
XX
PS Claim 1; SEQ ID NO 74; 186pp; English.
XX
CC The invention describes a method of identifying (M1) a compound to treat
CC a neuropsychiatric disorder. The method involves contacting a cell with a
CC test compound, determining the expression of one or more signature genes
CC by the cell, each signature gene comprising a nucleic acid that
CC hybridises to a nucleic acid chosen from 154 fully defined sequences (S1)
CC of e.g.: 4154, 1582, 645, 367, 405, 1080, 580, 1159, 539, 463, 2002,
CC 2469, 3165, 421, 2957, 644, 2562, 2858, 1194 and 1227 base pairs as given
CC in the specification and their complements, and comparing determined
CC expression of one or more signature genes to expression in cell not
CC contacted with test compound. Also described are: selecting (M2) one or
CC more signature genes that are indicative of an effective therapy for
CC treating a neuropsychiatric disorder; and a kit (I) for detecting an
CC electroconvulsive seizure (ECS) gene signature, comprising several
CC oligonucleotides, each of which is capable of specifically hybridising to
CC a different ECS signature gene. Therapeutic methods, which use ECS gene
CC signatures and ECS signature genes to treat, ameliorate or prevent
CC neuropsychiatric disorders, is also disclosed. (M1) is useful for
CC identifying a compound to treat a neuropsychiatric disorder chosen from
CC schizophrenia, autism, major depressive disorder (MDD), bipolar affective
CC disorder (BAD) and psychotic depression. (M1) is useful for identifying
CC compounds for treating neuropsychiatric disorders such as anxiety
CC disorders, eating disorders and attention deficit hyperactivity disorder.
CC The ECS signature genes of (M1) are useful in diagnostic and prognostic
CC methods of identifying people who have or who are susceptible to
CC neuropsychiatric disorders such as MDD and BAD. (M1) enables multi-
CC parameter high throughput screening of compounds for treating
CC neuropsychiatric disorders; screening of large number of candidate
CC compounds for treating neuropsychiatric disorder and enables cost-
CC effective and rapid screening of compounds useful as neuropsychiatric
CC drugs. This sequence represents an electroconvulsive seizure (ECS)
CC signature gene.
XX
SQ Sequence 4562 BP; 981 A; 1273 C; 1195 G; 1113 T; 0 U; 0 Other;

Query Match 50.4%; Score 25.2; DB 13; Length 4562;
Best Local Similarity 78.9%; Pred. No. 43;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTACCCACCCTGGATCTTGGCCCGCCCTTGAATTC 38
DB 3178 CCTATCGTGGCTGGATCTTGGCCCGCCACACCTTGAGTTC 3215

RESULT 13
ADD47107/c
ID ADD47107 standard; DNA; 28564 BP.
XX
AC ADD47107;

```


CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (described in Table 3 of
CC the specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 28564 BP; 7870 A; 6197 C; 6094 G; 7880 T; 0 U; 523 Other;

Query Match 50.4%; Score 25.2; DB 10; Length 28564;
Best Local Similarity 71.7%; Pred.No.53;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 CCTACCACCTGGATCCTTGTGCCCGCCCTTGAAATTCACAACCT 46
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21050 CCTATCCACCTCCCTCTATTCCCCCCCCTGACTTCCCCTACCT 21005

RESULT 16
ADD47111/c
ID ADD47111 standard; DNA; 28564 BP.
XX
AC ADD47111;
XX
XX
DT 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX
XX
DE Rat gene M19359, SEQ ID NO 12801.
XX
KW Rat; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
XX Rattus norvegicus.
OS Unidentified.
OS
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-033347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI
XX WFI; 2003-268312/26.
DR GENBANK; M19359.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Example 1; Page; 1017pp; English.
PS
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the

CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 28564 BP; 7870 A; 6197 C; 6094 G; 7880 T; 0 U; 523 Other;

Query Match 50.4%; Score 25.2; DB 10; Length 28564;
Best Local Similarity 71.7%; Pred. No. 53;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CCTACCCACCCCTGGATCTTGCCTCCGCGCCCTTGAATTCCTCCACCCCT 46
DB 21050 CCTATCCACCCCTCCCTCTATTCCTCCGCGCCCTTGAATTCCTCCACCT 21005

RESULT 19

ADBS7554/c
ID ADBS7554 standard; DNA; 28564 BP.

XX
AC ADBS7554;

XX 29-JAN-2004 (first entry)

XX Rat gene M19359, SEQ ID NO 3416.

DE Rat; ds; gene; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

OS Rattus norvegicus.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-033347P.

XX (GENO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

DR GENBANK; M19359.

XX

PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 28564 BP; 7870 A; 6197 C; 6094 G; 7880 T; 0 U; 523 Other;

Query Match 50.4%; Score 25.2; DB 10; Length 28564;
Best Local Similarity 71.7%; Pred. No. 53;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CCTACCCACCCCTGGATCTTGCCTCCGCGCCCTTGAATTCCTCCACCCCT 46

DB 21050 CCTATCCACCCCTCCCTCTATTCCTCCGCGCCCTTGAATTCCTCCACCT 21005

RESULT 20

ACN44110/c

ID ACN44110 standard; DNA; 223556 BP.

XX ACN44110;

XX 18-NOV-2004 (first entry)

XX Human genomic sequence HCG38337.

DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

KW Homo sapiens.

OS WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX

XX PF 02-OCT-2001; 2001WO-US030589.
XX XX
XX PR 02-OCT-2000; 2000US-0237054P.
XX XX
XX PA (GENE-) GENE LOGIC INC.
XX XX
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX XX
XX DR WPI; 2002-426119/45.
XX XX
XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular
XX PT carcinoma or metastatic liver tumor in a patient, involves detecting the
XX PT level of expression of two or more genes in a liver tissue sample.
XX XX
XX PS Claim 1; SEQ ID NO 1199; 298pp; English.
XX XX
XX CC The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX CC for this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX XX
XX SQ Sequence 211 BP; 43 A; 60 C; 55 G; 53 T; 0 U; 0 Other;

Query Match 45.6%; Score 22.8; DB 6; Length 211;
Best Local Similarity 79.4%; Pred. No. 2.1e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 17 CCTTGGCCCGCCCTTGAAATTCACACCTCCAC 50
DB 74 CCTTGGCCACCCCTTGCTCTCCACCCCCAC 107

RESULT 34
ABQ21641
ID ABQ21641 standard; DNA; 655 BP.
XX AC ABQ21641;
XX XX
XX DT 12-JUL-2002 (first entry)
XX XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 8232.
XX XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX PN WO200218632-A2.
XX PD 07-MAR-2002.
XX XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX PR 05-SEP-2000; 2000DE-01044543.
XX XX
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX DR WPI; 2002-371829/40.
XX XX

XX DR WPI; 2002-371829/40.
XX XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
XX PT from chemically treated DNA.
XX XX
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX XX
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one member,
XX CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX CC degree of hybridisation to both classes is determined from the label on
XX CC the amplicon. From the ratio of labels hybridised to the two classes of
XX CC oligomers, the degree of methylation is calculated. The method is used:
XX CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX CC and of a wide range of diseases, e.g. cancer, disorders of the central
XX CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX CC particularly by detecting mutations or single nucleotide polymorphisms
XX CC (SNP's); and (ii) for differentiation of cell or tissue types and for
XX CC investigating cell differentiation. The method allows the methylation
XX CC status of many C residues to be determined simultaneously. ABQ13410-
XX CC ABQ54121 represent genomic DNA sequences used to illustrate the method
XX CC for determining the degree of cytosine methylation described in the
XX CC disclosure of the invention
XX XX
XX SQ Sequence 655 BP; 276 A; 149 C; 36 G; 194 T; 0 U; 0 Other;

Query Match 45.6%; Score 22.8; DB 6; Length 655;
Best Local Similarity 66.0%; Pred. No. 2.4e+02;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCCCTGGATCTTGCCCGCCCTTGAAATTCACACCTCCAC 50
DB 593 COTCCGCCACCCCTACCTCCACCCCGCCCTACCTCCCGCCCGCCCTAC 642

RESULT 35
ABQ21640/c
ID ABQ21640 standard; DNA; 655 BP.
XX AC ABQ21640;
XX XX
XX DT 12-JUL-2002 (first entry)
XX XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 8231.
XX XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX PN WO200218632-A2.
XX PD 07-MAR-2002.
XX XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX PR 05-SEP-2000; 2000DE-01044543.
XX XX
XX PA (EPIG-) EPIGENOMICS AG.
XX XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX DR WPI; 2002-371829/40.
XX XX

PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

XX Sequence 655 BP; 194 A; 36 C; 149 G; 276 T; 0 U; 0 Other;

Query Match 45.6%; Score 22.8; DB 6; Length 655;

Best Local Similarity 66.0%; Pred. No. 2.4e+02;

Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTATCCACCCGATCTTCGCGCCGCGCCCTTGATTCGACCCCTCCAC 50

DB 63 CCTCCACCCCTACCTCCGACCCGCGCCCTACCTCCGACCCGCGCCCTAC 14

RESULT 36

AL04333/c

ID AAL04333 standard; DNA; 945 BP.

AC AAL04333;

DT 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 7021.

KW Human; reproductive system related antigen; reproductive system disorder;

KW cancer; gene therapy; ds.

OS Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.


```

XX AAV99094;
AC 16-MAR-1999 (first entry)
DT
DE DNA methyltransferase genomic fragment.
XX
XX DNA methyltransferase; DNA Mefase; antisense oligonucleotide; human;
KW cellular growth; tumour growth inhibition; silenced gene activation;
KW beta thalassemia; sickle cell anemia; ss.
XX
XX Homo sapiens.
OS
XX WO9854313-A2.
PN
XX 03-DEC-1998.
PD
XX
XX 29-MAY-1998; 98WO-IB001107.
PF
XX
XX 30-MAY-1997; 97US-00866340.
PR
XX 17-DEC-1997; 97US-0069865P.
PR
XX (UYMC-) UNIV MCGILL.
PA
XX
XX Szyf M, Bigey P, Ramchandani S;
PI
XX WPI; 1999-059833/05.
DR
XX
XX New DNA methyltransferase nucleotide sequences - used particularly to
PT develop antisense oligonucleotides for diagnostic and therapeutic
PT purposes, particularly for inhibiting tumour growth.
XX
XX Claim 1; Page 38-40; 108pp; English.
PS
XX
XX AAV99091-128 represent DNA methyltransferase (DNA Mefase) genomic
CC fragments. The sequences together sequentially produce the full length
CC sequence AAV99128. The sequences can be used to derive antisense
CC oligonucleotides which inhibit DNA Mefase expression. The antisense
CC oligonucleotides can be used in investigating the role of DNA Mefase in
CC cellular growth. They can be administered at different points in the cell
CC cycle, or in conjugation with promoters or inhibitors of cell growth to
CC determine the role of DNA Mefase in the growth of the cell type of
CC interest. The antisense oligonucleotides can also be used for inhibiting
CC tumour growth in a mammal, or to activate silenced genes to provide a
CC missing gene function. This ameliorates disease symptoms, e.g. in beta
CC thalassemia and sickle cell anemia. The antisense oligonucleotides can
CC also be used as analytical and diagnostic tools and a potentiators of
CC transgenic plant and animal studies
XX
XX Sequence 4459 BP; 1166 A; 1076 C; 923 G; 1294 T; 0 U; 0 Other;
SQ
Query Match 45.6%; Score 22.8; DB 2; Length 4459;
Best Local Similarity 66.0%; Pred. No. 2.9e+02;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 CCTACCCACCTGGATCCTTGGCCGCCCTTGAATTCCTCCACCTCCAC 50
DB 380 CTTGCTTCCCTGGAGCTTGACTCCCAATTCATCTCTCAAAATGCCAC 429
RESULT 39
AAA95949/c
ID AAA95949 standard; DNA; 8280 BP.
XX
AC AAA95949;
XX
XX 02-FEB-2001 (first entry)
DT
XX Human KLK-L6 gene.
DE
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
KW kallikrein-like protein; serine protease; cytotatic; cancer;
KW prostrate cancer; ds.

```

```

XX Homo sapiens.
OS
XX WO200053776-A2.
PN
XX 14-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000WO-CA000258.
PF
XX
XX 11-MAR-1999; 99US-0124260P.
PR
XX 01-APR-1999; 99US-0127386P.
PR
XX 21-JUL-1999; 99US-0144919P.
PR
XX (MOUN ) MOUNT SINAI HOSPITAL.
PA
XX
XX Yousef GM, Diamandis EP;
PI
XX WPI; 2000-587440/55.
DR
XX P-PSDB; AAB21305, AAB21306.
XX
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer.
PT
XX Claim 1; Page 174-178; 184pp; English.
PS
XX
XX The present sequence is the long form of the human KLK-L6 gene, which
CC encodes a kallikrein-like protein. Kallikreins and kallikrein-like
CC proteins are a subgroup of the serine protease enzyme family. They
CC catalyse the selective cleavage of specific polypeptide precursors to
CC release peptides with potent biological activity. Nucleic acids encoding
CC kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5 and KLK-
CC L6 have been isolated. The proteins are useful in the treatment,
CC monitoring and diagnosis of cancers, especially prostate cancer. They
CC can also be used to identify a substance that can associate with or
CC mediate the biological activity of the proteins. Antibodies can be used
CC to treat conditions mediated by the kallikrein-like proteins
XX
XX Sequence 8280 BP; 1804 A; 2392 C; 2246 G; 1838 T; 0 U; 0 Other;
SQ
Query Match 45.6%; Score 22.8; DB 3; Length 8280;
Best Local Similarity 71.4%; Pred. No. 3.2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 CCTACCCACCTGGATCCTTGGCCGCCCTTGAATTCCTCCAA 42
DB 900 CTTACCTCTCTCAGCCCTCCCTAGCCCTTAAATTCCTCCA 859
RESULT 40
ADN10931/c
ID ADN10931 standard; DNA; 8280 BP.
XX
AC ADN10931;
XX
XX 01-JUL-2004 (first entry)
DT
XX Human kallikrein 14 gene, marker of endocrine cancer.
DE
XX Human; kallikrein 14; endocrine cancer; ovarian cancer; marker;
KW diagnosis; gene; ds.
KW
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH misc_signal 784..799
FT /*tag= a
FT /note= "Putative androgen response element"
FT CDS 2803..7410
FT /*tag= c
FT /product= "Human kallikrein 14"
FT exon 2803..2842
FT /*tag= b
FT intron 2843..3713

```


PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250160P.
PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232400P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234274P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234997P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239935P.	XX		
PR	13-OCT-2000;	2000US-0239937P.	XX		
PR	20-OCT-2000;	2000US-0240960P.	XX		
PR	20-OCT-2000;	2000US-0241221P.	XX		
PR	20-OCT-2000;	2000US-0241785P.	XX		
PR	20-OCT-2000;	2000US-0241786P.	XX		
PR	20-OCT-2000;	2000US-0241787P.	XX		
PR	20-OCT-2000;	2000US-0241808P.	XX		
PR	20-OCT-2000;	2000US-0241809P.	XX		
PR	01-NOV-2000;	2000US-0244617P.	XX		
PR	08-NOV-2000;	2000US-0246474P.	XX		
PR	08-NOV-2000;	2000US-0246475P.	XX		
PR	08-NOV-2000;	2000US-0246476P.	XX		
PR	08-NOV-2000;	2000US-0246477P.	XX		
PR	08-NOV-2000;	2000US-0246478P.	XX		
PR	08-NOV-2000;	2000US-0246523P.	XX		
PR	08-NOV-2000;	2000US-0246524P.	XX		
PR	08-NOV-2000;	2000US-0246525P.	XX		
PR	08-NOV-2000;	2000US-0246526P.	XX		
PR	08-NOV-2000;	2000US-0246527P.	XX		
PR	08-NOV-2000;	2000US-0246528P.	XX		
PR	08-NOV-2000;	2000US-0246532P.	XX		
PR	08-NOV-2000;	2000US-0246609P.	XX		
PR	08-NOV-2000;	2000US-0246610P.	XX		
PR	08-NOV-2000;	2000US-0246611P.	XX		
PR	08-NOV-2000;	2000US-0246613P.	XX		
PR	17-NOV-2000;	2000US-0249207P.	XX		
PR	17-NOV-2000;	2000US-0249208P.	XX		
PR	17-NOV-2000;	2000US-0249209P.	XX		
PR	17-NOV-2000;	2000US-0249210P.	XX		
PR	17-NOV-2000;	2000US-0249211P.	XX		
PR	17-NOV-2000;	2000US-0249212P.	XX		
PR	17-NOV-2000;	2000US-0249213P.	XX		
PR	17-NOV-2000;	2000US-0249214P.	XX		
PR	17-NOV-2000;	2000US-0249215P.	XX		
PR	17-NOV-2000;	2000US-0249216P.	XX		
PR	17-NOV-2000;	2000US-0249217P.	XX		
PR	17-NOV-2000;	2000US-0249218P.	XX		
PR	17-NOV-2000;	2000US-0249			

misc_feature	7556..7557	/*tag= e	/note= "approximately 1500 unspecified nucleotides are present between these bases"
misc_feature	7866..7867	/*tag= f	/note= "approximately 700 unspecified nucleotides are present between these bases"
misc_feature	8542..8543	/*tag= g	/note= "approximately 1000 unspecified nucleotides are present between these bases"
misc_feature	8807..8808	/*tag= h	/note= "approximately 1700 unspecified nucleotides are present between these bases"
misc_feature	9099..9100	/*tag= i	/note= "approximately 450 unspecified nucleotides are present between these bases"
misc_feature	9355..9356	/*tag= j	/note= "these nucleotides are not present in AAV991103"
misc_feature	9662..9663	/*tag= k	/note= "350 unspecified nucleotides are present between these bases"
misc_feature	11407..11409	/*tag= l	/note= "300 unspecified nucleotides are present between these bases"
misc_feature	11704..11705	/*tag= m	/note= "approximately 450 unspecified nucleotides are present between these bases"
misc_feature	13274..13323	/*tag= n	/note= "approximately 900 unspecified nucleotides are present between these bases"
misc_feature	13049..13050	/*tag= o	/note= "approximately 180 unspecified nucleotides are present between these bases"
misc_feature	13321..13322	/*tag= p	/note= "these nucleotides as stated as being from AAV99111 in the figure, but are actually from AAV99112"
misc_feature	13328	/*tag= q	/note= "approximately 1500 unspecified nucleotides are present between these bases"
misc_feature	13566..13567	/*tag= r	/note= "this nucleotide is not present in AAV99112"
misc_feature	13827..13828	/*tag= s	/note= "approximately 150 unspecified nucleotides are present between these bases"
misc_feature	14073..14074	/*tag= t	/note= "400 unspecified nucleotides are present between these bases"
misc_feature	14359..14360	/*tag= u	/note= "approximately 850 unspecified nucleotides are present between these bases"
misc_feature	14796..14797	/*tag= v	/note= "approximately 450 unspecified nucleotides are present between these bases"
misc_feature		/*tag= w	/note= "approximately 520 unspecified nucleotides are present between these bases"

FT	misc_feature	present between these bases"
FT		15194..15195
FT		/tag= x
FT		/note= "approximately 950 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	15869..15870
FT		/tag= y
FT		/note= "approximately 500 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	16324..16325
FT		/tag= z
FT		/note= "approximately 250 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	16764..16765
FT		/tag= aa
FT		/note= "approximately 1200 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	17085..17086
FT		/tag= ab
FT		/note= "approximately 600 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	17578..17579
FT		/tag= ac
FT		/note= "approximately 600 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	17868..17869
FT		/tag= ad
FT		/note= "approximately 800 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	18254..18253
FT		/tag= ae
FT		/note= "approximately 400 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	18611..18612
FT		/tag= af
FT		/note= "approximately 1450 unspecified nucleotides are
FT		present between these bases"
FT	misc_feature	18869..18870
FT		/tag= ag
FT		/note= "approximately 550 unspecified nucleotides are
FT		present between these bases"
XX		
XX	W09854313-A2.	
XX		
PD	03-DEC-1998.	
XX		
PF	29-MAY-1998;	98WO-IB001107.
XX		
PR	30-MAY-1997;	97US-00866340.
PR	17-DEC-1997;	97US-0069865P.
XX		
PA	(UYMC)- UNIV MCGILL.	
XX		
PI	Szyf M, Bigey P, Ramchandani S;	
XX		
DR	WPI; 1999-059833/05.	
XX		
XX		
PT	New DNA methyltransferase nucleotide sequences - used particularly to	
PT	develop antisense oligonucleotides for diagnostic and therapeutic	
PT	purposes, particularly for inhibiting tumour growth.	
XX		
PS	Claim 1; Fig 1; 108pp; English.	
XX		
CC	The present sequence represents a full length DNA methyltransferase (DN	
CC	MeTase) genomic sequence. It is derived from fragments ANV99091-128. Th	
CC	sequences can be used to derive antisense oligonucleotides which inhibi	
CC	DNA MeTase expression. The antisense oligonucleotides can be used in	
CC	investigating the role of DNA MeTase in cellular growth. They can be	
CC	administered at different points in the cell cycle, or in conjugation	
CC	with promoters or inhibitors of cell growth to determine the role of DN	
CC	MeTase in the growth of the cell type of interest. The antisense	
CC	oligonucleotides can also be used for inhibiting tumour growth in a	
CC	mammal, or to activate silenced genes to provide a missing gene function	

XX DE Human adult heart cDNA #2204.
XX KW Human; es; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX DR WPI; 2003-615964/58.
XX PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX PS Claim 1; SEQ ID NO 5102; 44pp; English.
XX CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX SQ Sequence 435 BP; 85 A; 166 C; 107 G; 77 T; 0 U; 0 Other;
Query Match 45.2%; Score 22.6; DB 9; Length 435;
Best Local Similarity 68.9%; Pred. NO. 2.7e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 4 ACCACCCCTGGATCCTTGGCCCGCCCTTGAATTCACACCTCC 48
DB 156 ACACACCCCTGGACCTTGGCCCACTCTCTCCCGCCCGCCCTCC 200
RESULT 46
ADQ52701/c
ID ADQ52701 standard; DNA; 557 BP.
XX AC ADQ52701;
XX 21-OCT-2004 (first entry)
XX DE Novel canine microarray-related DNA sequence SeqID4003.
XX KW canine microarray; drug screening; toxicity assay;

KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
XX protein adduct formation; hepatitis; dog; de.
XX OS Canis familiaris.
XX PN WO2004063324-A2.
XX PD 29-JUL-2004.
XX PF 05-MAY-2003; 2003WO-US013853.
XX PR 03-MAY-2002; 2002US-0377240P.
XX PA (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX PI Diggans JC, Porter M, Wei T;
XX DR WPI; 2004-561890/54.
XX PT New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX PS Claim 1; SEQ ID NO 4003; 41pp; English.
XX CC This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northernblots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX SQ Sequence 557 BP; 99 A; 193 C; 128 G; 132 T; 0 U; 5 Other;
Query Match 45.2%; Score 22.6; DB 13; Length 557;
Best Local Similarity 68.9%; Pred. NO. 2.7e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCACACCC 45
DB 335 CCCACCCCGCTGGACCTTGGCCCACTCTCTCCCGCCCTCC 291
RESULT 47
AAK92116/c
ID AAK92116 standard; cDNA; 827 BP.
XX AC AAK92116;

XX WPI; 2004-204755/20.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX Example 18; SEQ ID NO 1838; 1340pp; English.
XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'
CC ends using the oligo-capping method. This polynucleotide sequence is the
CC 5' end of a representative human DNA cluster of the invention.
XX
XX Sequence 827 BP; 183 A; 250 C; 249 G; 139 T; 0 U; 6 Other;
SQ
Query Match 45.2%; Score 22.6; DB 12; Length 827;
Best Local Similarity 68.9%; Pred. No. 2.9e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 1 CCTACCCACCCCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCC 45
Db 724 CCAACCCCTTGGGTCTTGTCTCCCGCCACAGAGCCCGCAGCTC 680
RESULT 50
ADL28543/c
ID ADL28543 standard; cDNA; 827 BP.
XX
XX AC ADL28543;
XX 20-MAY-2004 (first entry)
XX 5' end of a human cDNA molecule SeqID 576.
XX human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method; ss.
XX Homo sapiens.
XX EP1396543-A2.
XX PN
XX PD 10-MAR-2004.
XX PF 07-JUL-2000; 2003EP-00025638.
XX PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Iehli S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX Disclosure; SEQ ID NO 576; 1340pp; English.
XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method

CC for efficiently cloning a full length human cDNA from both the 5' and 3'
CC ends using the oligo-capping method. This polynucleotide sequence is the
CC 5' end of a full length human cDNA sequence of the invention.
XX
XX Sequence 827 BP; 183 A; 250 C; 249 G; 139 T; 0 U; 6 Other;
SQ
Query Match 45.2%; Score 22.6; DB 12; Length 827;
Best Local Similarity 68.9%; Pred. No. 2.9e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 1 CCTACCCACCCCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCC 45
Db 724 CCAACCCCTTGGGTCTTGTCTCCCGCCACAGAGCCCGCAGCTC 680
Search completed: April 25, 2005, 11:36:20
Job time : 518.707 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:13:24 ; Search time 4302.59 Seconds
(without alignments)
442.342 Million cell updates/sec

Title: US-10-010-476-13
Perfect score: 50
Sequence: 1 CTTACCCACCCGGATCCTT.....TTGAATTCACACCTCCAC 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27.6	55.2	908	9	AG081706 Pan trogl
2	27	54.0	594	7	CO881510 BovGen.09
3	26.6	53.2	619	9	FR0006159
C 4	26.6	53.2	1009	9	CL503275 SAIL 721
5	26.4	52.8	1096	8	BZ558473
C 6	26.2	52.4	810	4	BG368058 HVSM1001
7	26.2	52.4	960	6	CA790467 AGENCOURT
8	26	52.0	288	1	AV166399 AV166399
9	26	52.0	365	7	CK838096 4063391 B
C 10	26	52.0	368	8	BH639368
11	26	52.0	386	6	BY605779
12	26	52.0	668	7	CK952167
13	26	52.0	699	1	AL574035
14	26	52.0	726	8	AQ253286 HS 2046 B
15	26	52.0	739	7	CK833690 4057753 B
16	26	52.0	877	1	AL522831
C 17	26	52.0	966	9	AG331887 Mus muscu
18	26	52.0	1057	4	BM470168 AGENCOURT
19	25.8	51.6	283	7	CF094505 QHN18C01
20	25.8	51.6	933	5	BO943145 AGENCOURT
C 21	25.6	51.2	309	9	CE491137 tigr-g88-
C 22	25.6	51.2	388	7	CNA38735 BE04017A1
23	25.6	51.2	444	5	BY356513
C 24	25.6	51.2	1164	5	BQ953457 AGENCOURT

98	24.4	48.8	1041	9	CL489692	CL489692 SAIL 528	171	23.8	47.6	474	2	AW742116	AW742116 up52809.Y
99	24.4	48.8	1065	8	CC233014	BQ261-132	C 172	23.8	47.6	479	2	BB852074	BB852074 uw15c07.x
100	24.4	48.8	1069	5	BQ411252	AGENCOURT	C 173	23.8	47.6	486	4	BQ606090	BQ606090 L0906F09-
101	24.4	48.8	1071	4	BM469279	AGENCOURT	174	23.8	47.6	488	6	CA320082	UI-M-FV0-
102	24.4	48.8	1078	8	CC244889	CH261-136	175	23.8	47.6	494	2	BB735534	BB735534
103	24.4	48.8	1082	8	CC272947	CH261-131	176	23.8	47.6	507	6	CD804126	UI-M-GV0-
104	24.4	48.8	1153	6	CA468223	AGENCOURT	177	23.8	47.6	512	5	BX511712	BX511712
105	24.4	48.8	1157	8	CC319166	TAM32-370	178	23.8	47.6	522	6	CD351362	UI-M-GI0-
106	24.4	48.8	1161	5	BM909256	AGENCOURT	C 179	23.8	47.6	523	2	AW537841	C0100B02-
107	24.4	48.8	1197	4	BM801411	AGENCOURT	C 180	23.8	47.6	527	1	AI646867	ub6907.x
108	24.4	48.8	1207	4	BM702897	AGENCOURT	C 181	23.8	47.6	529	5	BY477480	BY477480
109	24.4	48.8	1208	4	BM465604	AGENCOURT	C 182	23.8	47.6	530	2	BF466953	UI-M-CG0p
110	24.4	48.8	1366	4	BM459149	AGENCOURT	C 183	23.8	47.6	539	2	BB770233	BB770233
111	24.4	48.8	1389	8	CC321868	TAM32-34N	184	23.8	47.6	544	2	BB730755	BB730755
112	24.2	48.4	384	8	AQ292371	HS 2180.A	C 185	23.8	47.6	547	6	CA871634	K0911H06-
113	24.2	48.4	575	4	BG823936	602727224	C 186	23.8	47.6	548	4	BM204387	C0268F08-
114	24.2	48.4	694	7	CN852687	Ha_mx0_03	C 187	23.8	47.6	559	4	BI647566	603278145
115	24.2	48.4	694	9	AG139528	Pan trogl	C 188	23.8	47.6	561	2	AW543558	C0171C11-
116	24.2	48.4	711	7	W23214	80B5 Human	C 189	23.8	47.6	562	5	BQ830593	LL61n1087
117	24.2	48.4	780	9	AG520705	Mus muscu	C 190	23.8	47.6	563	9	CE821594	tigr-g89-
118	24.2	48.4	798	7	CN439366	BE04019A1	C 191	23.8	47.6	565	5	BQ101899	ih87f03.x
119	24.2	48.4	966	1	AL577035	AL577035	C 192	23.8	47.6	585	5	BF764144	BP764144
120	24.2	48.4	967	9	CNS018BK	AL109130 Drocephil	C 193	23.8	47.6	587	7	CK329455	H8165D06-
121	24.2	48.4	1478	9	AG031693	Pan trogl	C 194	23.8	47.6	590	6	CA881526	K0991B05-
122	24.2	48.4	352	7	CF119570	MTU10CS.P	C 195	23.8	47.6	594	6	CA879200	K0972E08-
123	24	48.0	360	2	BF465438	UI-M-CG0p	196	23.8	47.6	628	2	BF120928	601757737
124	24	48.0	619	9	CNS044RQY	AL304243 Tetraodon	197	23.8	47.6	643	9	CR023726	Forward 8
125	24	48.0	662	7	CK846486	968990 MA	198	23.8	47.6	662	8	AG509226	nbx000941
126	24	48.0	723	9	AG381451	Mus muscu	199	23.8	47.6	663	9	AG299700	Mus muscu
127	24	48.0	772	9	AG517523	Mus muscu	200	23.8	47.6	666	6	CD803106	UI-M-GV0-
128	24	48.0	792	6	CA088972	SCRFA213	201	23.8	47.6	746	6	CA326358	UI-M-F20-
129	24	48.0	854	2	BE878567	601492811	202	23.8	47.6	775	5	BU560911	AGENCOURT
130	24	48.0	875	4	BG344078	HVME9000	203	23.8	47.6	933	5	BQ651947	AGENCOURT
131	24	48.0	905	5	BU543312	AGENCOURT	204	23.8	47.6	965	5	BQ900707	AGENCOURT
132	24	48.0	928	6	CA460791	AGENCOURT	205	23.8	47.6	1110	4	BM472930	AGENCOURT
133	24	48.0	947	5	BQ420627	AGENCOURT	206	23.8	47.6	2763	3	AK028235	Mus muscu
134	24	48.0	988	9	AG377177	Mus muscu	207	23.6	47.2	180	9	CR255716	Reverse 8
135	24	48.0	1039	5	BQ924078	AGENCOURT	C 208	23.6	47.2	278	6	CB482841	ins99 A03
136	24	48.0	1040	6	CA900854	AGENCOURT	C 209	23.6	47.2	289	1	AV014365	AV014365
137	24	48.0	1101	1	AL071156	Drosophil	C 210	23.6	47.2	334	9	CR171758	Forward 8
138	24	48.0	1126	4	BM802967	AGENCOURT	211	23.6	47.2	336	2	BE217652	BB217652
139	24	48.0	1158	5	BX898637	BX898637	212	23.6	47.2	352	9	CE615781	W61990 md83b02.r1
140	24	48.0	1162	9	CL030896	CH216-31L	C 213	23.6	47.2	373	7	BI961920	MON01 8.D
141	24	48.0	1168	5	BM907604	AGENCOURT	C 214	23.6	47.2	413	4	BI961920	me50b02.r1
142	24	48.0	1206	8	CC253696	CH261-150	C 215	23.6	47.2	417	7	W75418	W75418
143	24	48.0	1537	8	CC237106	CH261-191	C 216	23.6	47.2	421	5	BQ760688	BQ760688
144	23.8	47.6	260	4	BM118623	L0913F08-	217	23.6	47.2	424	1	AI653867	ty02e03.x
145	23.8	47.6	261	4	BM199190	C0454G11-	218	23.6	47.2	471	8	AQ165167	HS 3022.B
146	23.8	47.6	286	5	BP769962	BP769962	219	23.6	47.2	496	1	AJ708828	AJ708828
147	23.8	47.6	287	4	BM207345	C0611E03-	C 220	23.6	47.2	511	2	B283460	BB283460
148	23.8	47.6	291	1	AI465559	VW78F01.Y	C 221	23.6	47.2	517	4	BQ067160	H3051C07-
149	23.8	47.6	302	4	BM203900	C0261B04-	C 222	23.6	47.2	520	6	CA884170	B0108D11-
150	23.8	47.6	305	5	BQ563614	gi05a04.Y	223	23.6	47.2	560	1	AI225464	ue87n11.Y
151	23.8	47.6	312	4	BG147226	mab50b02.	224	23.6	47.2	568	8	AZ501993	AZ501993
152	23.8	47.6	314	5	BQ563613	gi05a04.x	225	23.6	47.2	576	2	BE289377	601089353
153	23.8	47.6	324	2	AW742752	UD52d03.x	C 226	23.6	47.2	624	4	BJ481546	BJ481546
154	23.8	47.6	333	4	BM213545	C0838A07-	C 227	23.6	47.2	644	2	BE003815	BB003815
155	23.8	47.6	375	1	AL118319	V9111B46	C 228	23.6	47.2	653	8	AZ121254	RPC1-23-1
156	23.8	47.6	375	1	AU255786	AU255786	229	23.6	47.2	674	6	BY757487	BY757487
157	23.8	47.6	388	5	BY154437	BY154437	230	23.6	47.2	688	9	CR047114	Forward 8
158	23.8	47.6	390	6	BY657286	BY657286	231	23.6	47.2	694	7	CK948082	CK948082
159	23.8	47.6	405	1	AA838894	VW78F01.X	C 232	23.6	47.2	698	6	CD469341	LeukoS2.3
160	23.8	47.6	410	1	AI157998	ud22e08.r	C 233	23.6	47.2	702	8	AZ260291	RPC1-23-1
161	23.8	47.6	412	5	BM989850	UI-M-DJ1-	234	23.6	47.2	706	9	AG098722	Pan trogl
162	23.8	47.6	415	2	AW123590	UI-M-BH2.	C 235	23.6	47.2	763	8	B2274799	BB274799
163	23.8	47.6	437	1	AU040891	AU040891	C 236	23.6	47.2	782	2	BF627244	HVMEB000
164	23.8	47.6	444	1	AU040890	AU040890	237	23.6	47.2	849	9	CC913661	t079m18ba
165	23.8	47.6	450	2	AW123412	UI-M-BH2.	238	23.6	47.2	852	9	AG288036	Mus muscu
166	23.8	47.6	456	1	AI837779	UI-M-AK0-	C 239	23.6	47.2	877	9	BX959427	Forward 8
167	23.8	47.6	456	2	AW909566	ur76h11.x	240	23.6	47.2	890	4	BI525356	602824465
168	23.8	47.6	459	2	BB696794	BB696794	241	23.6	47.2	928	5	BQ935572	AGENCOURT
169	23.8	47.6	462	1	AI845993	UI-M-API-	242	23.6	47.2	958	9	CL475669	SAIL 242
170	23.8	47.6	467	2	BB747953	BB747953	243	23.6	47.2	987	9	AG354127	Mus muscu

C 244	23.6	47.2	1122	9	AG424409	Mus muscu	317	23.2	46.4	374	2	BB804069	BB804069
C 245	23.6	47.2	1134	5	BQ331762	AGENCOURT	318	23.2	46.4	377	2	BB808613	BB808613
C 246	23.6	47.2	1193	5	BQ677826	AGENCOURT	319	23.2	46.4	395	5	BY451970	BY451970
C 247	23.6	47.2	1235	6	CA456292	AGENCOURT	320	23.2	46.4	407	5	BQ932102	AGENCOURT
C 248	23.6	47.2	1263	6	CA494360	AGENCOURT	321	23.2	46.4	416	2	BB800119	BB800119
C 249	23.6	47.2	1284	8	CC236768	CH261-141	322	23.2	46.4	432	5	BY363763	BY363763
C 250	23.6	47.2	1398	3	AK028795	Mus muscu	323	23.2	46.4	442	7	CN963091	10588_100
C 251	23.4	46.8	210	9	CL437805	AGENCOURT	324	23.2	46.4	507	2	BB753706	BB753706
C 252	23.4	46.8	271	2	BB608158	BB608158	C 325	23.2	46.4	509	2	AW493980	UT-M-BH3-
C 253	23.4	46.8	349	2	BE229719	98SN0107	C 326	23.2	46.4	532	2	BF628269	HVSMED000
C 254	23.4	46.8	392	2	BF902053	IL5-WT021	C 327	23.2	46.4	540	5	BY412408	BY412408
C 255	23.4	46.8	405	7	N49634	YV25C12.r1	C 328	23.2	46.4	548	4	BJ643218	BJ643218
C 256	23.4	46.8	408	7	T80711	YD23F03.r1	C 329	23.2	46.4	576	5	BP768096	BP768096
C 257	23.4	46.8	438	9	CR088925	Forward s	C 330	23.2	46.4	578	2	BB276786	BB276786
C 258	23.4	46.8	461	9	CL356120	RPCI-144_41	C 331	23.2	46.4	581	5	BF759790	BF759790
C 259	23.4	46.8	495	8	BM107128	RPCI-24-2	C 332	23.2	46.4	622	2	BF795431	BF795431
C 260	23.4	46.8	498	6	CB256740	62-E01166	C 333	23.2	46.4	638	4	BM359716	GA_Ea002
C 261	23.4	46.8	507	2	AW658813	95560 MAR	C 334	23.2	46.4	639	5	BP767354	BP767354
C 262	23.4	46.8	561	1	AV526534	AV526534	C 335	23.2	46.4	647	6	BY723700	BY723700
C 263	23.4	46.8	575	6	CB449591	703802 MA	C 336	23.2	46.4	655	2	BB794659	BB794659
C 264	23.4	46.8	576	9	AG576446	Mus muscu	C 337	23.2	46.4	662	7	CO044177	UT-M-GH0-
C 265	23.4	46.8	578	6	CD816131	BN15_028N	C 338	23.2	46.4	668	2	BB556663	BB556663
C 266	23.4	46.8	579	6	CD815424	BN15_026F	C 339	23.2	46.4	675	2	BB794609	BB794609
C 267	23.4	46.8	589	6	CD813617	BN15_020F	C 340	23.2	46.4	691	2	BB643723	BB643723
C 268	23.4	46.8	601	9	CE445603	t1gr-988-	C 341	23.2	46.4	697	4	BI463324	603204339
C 269	23.4	46.8	617	6	CD816118	BN15_028U	C 342	23.2	46.4	748	9	AG534886	Mus muscu
C 270	23.4	46.8	640	1	AV822331	AV822331	C 343	23.2	46.4	779	9	AG170748	Pan trogl
C 271	23.4	46.8	668	8	AZ104538	RPCI-23-3	C 344	23.2	46.4	783	9	AG553739	Mus muscu
C 272	23.4	46.8	674	6	CD814643	BN15_023L	C 345	23.2	46.4	785	9	AG530474	Mus muscu
C 273	23.4	46.8	675	1	AA541878	VJ54H03.r	C 346	23.2	46.4	790	8	AZ684641	ENTHR61TR
C 274	23.4	46.8	709	6	CD836033	BN45_0471	C 347	23.2	46.4	809	9	AG184632	Pan trogl
C 275	23.4	46.8	717	9	CL158139	104_346.1	C 348	23.2	46.4	813	9	AG526931	Mus muscu
C 276	23.4	46.8	717	9	CL158140	104_346.1	C 349	23.2	46.4	830	6	CA219294	SCRPA111
C 277	23.4	46.8	726	6	CD815126	BN15_025F	C 350	23.2	46.4	839	6	CA461326	AGENCOURT
C 278	23.4	46.8	751	9	AG173190	Pan trogl	C 351	23.2	46.4	873	7	CF886134	tr1c085X0
C 279	23.4	46.8	762	7	CF870044	tr1c021x1	C 352	23.2	46.4	877	5	BU940779	AGENCOURT
C 280	23.4	46.8	763	6	CD836851	BN45_050C	C 353	23.2	46.4	914	5	BQ708833	AGENCOURT
C 281	23.4	46.8	806	9	CG137563	PUIDF277B	C 354	23.2	46.4	938	6	CB202503	AGENCOURT
C 282	23.4	46.8	813	6	CB958120	AGENCOURT	C 355	23.2	46.4	945	5	BU506088	AGENCOURT
C 283	23.4	46.8	820	5	BM985972	7_F03_T3	C 356	23.2	46.4	968	1	AL582734	AL582734
C 284	23.4	46.8	823	2	BF180509	601804305	C 357	23.2	46.4	978	5	BQ936704	AGENCOURT
C 285	23.4	46.8	830	2	BF180509	601804305	C 358	23.2	46.4	1003	9	CNS025WA	Tetrarodon
C 286	23.4	46.8	830	6	CB900205	tr1c021x1	C 359	23.2	46.4	1027	7	CK149172	1742HFEX1
C 287	23.4	46.8	851	6	CD572875	PBL 22_F0	C 360	23.2	46.4	1034	1	AL576069	AL576069
C 288	23.4	46.8	852	7	CF652323	48-I02036	C 361	23.2	46.4	1034	5	BQ066123	AGENCOURT
C 289	23.4	46.8	867	8	CC207920	CH261-108	C 362	23.2	46.4	1065	4	BI913778	603180743
C 290	23.4	46.8	886	9	CL464613	SAIL_1223	C 363	23.2	46.4	1121	5	BM917944	AGENCOURT
C 291	23.4	46.8	888	5	BQ721117	AGENCOURT	C 364	23.2	46.4	1123	5	BQ053473	AGENCOURT
C 292	23.4	46.8	893	2	BF239876	601905061	C 365	23.2	46.4	1184	4	BM468182	AGENCOURT
C 293	23.4	46.8	893	9	AG327046	Mus muscu	C 366	23.2	46.4	1222	8	CC308030	TAM32-14K
C 294	23.4	46.8	900	9	AG578851	Mus muscu	C 367	23.2	46.4	1274	5	BQ948175	AGENCOURT
C 295	23.4	46.8	935	5	AG541173	Mus muscu	C 368	23.2	46.4	1278	4	BM474556	AGENCOURT
C 296	23.4	46.8	936	5	BX421063	EX421063	C 369	23.2	46.4	1293	8	CC315260	TAM32-30E
C 297	23.4	46.8	942	9	CL069251	CH216-116	C 370	23.2	46.4	1298	3	AK012129	Mus muscu
C 298	23.4	46.8	955	9	CL491742	SAIL 55_B	C 371	23.2	46.4	1341	3	AK044193	Mus muscu
C 299	23.4	46.8	980	7	CF879980	tr1c021x1	C 372	23.2	46.4	1351	3	AK087819	Mus muscu
C 300	23.4	46.8	1011	9	CL493074	SAIL 573	C 373	23.2	46.4	1351	2	AA654053	nt81805_8
C 301	23.4	46.8	1022	9	AG396491	Mus muscu	C 374	23.2	46.4	1351	2	BB312516	BB312516
C 302	23.4	46.8	1023	9	CL144793	ISB1-144P	C 375	23.2	46.4	1351	2	BB545139	BB545139
C 303	23.4	46.8	1027	9	CNS000ZKN	drosophil	C 376	23.2	46.4	1351	2	BB604273	BB604273
C 304	23.4	46.8	1029	5	BM911311	AGENCOURT	C 377	23.2	46.4	1351	5	BY362888	BY362888
C 305	23.4	46.8	1029	8	CC275513	CH261-133	C 378	23.2	46.4	1351	7	H77172	17603_Lambd
C 306	23.4	46.8	1047	9	AG396402	Mus muscu	C 379	23.2	46.4	1351	1	AA292844	zt64d10_r
C 307	23.4	46.8	1059	5	CL091837	ISB1-21B1	C 380	23.2	46.4	1351	2	AW658183	94016 MAR
C 308	23.4	46.8	1066	5	BQ939555	AGENCOURT	C 381	23.2	46.4	1351	2	AA398662	zt64d10_8
C 309	23.4	46.8	1066	8	CC210194	CH261-161	C 382	23.2	46.4	1351	5	EX103001	EX103001
C 310	23.4	46.8	1109	4	BM470312	AGENCOURT	C 383	23.2	46.4	1351	2	CR803330	GR0AAA24A
C 311	23.4	46.8	1141	9	CL510440	SAIL 82_G	C 384	23.2	46.4	1351	1	AA536080	ntF95608_8
C 312	23.4	46.8	1276	5	BQ064118	AGENCOURT	C 385	23.2	46.4	1351	1	CR843602	GR0AAA78C
C 313	23.4	46.8	1305	8	CC291133	CH261-107	C 386	23.2	46.4	1351	6	CD003768	VVA013E02
C 314	23.4	46.8	1314	8	CC188378	CH261-129	C 387	23.2	46.4	1351	6	CB915119	VVD051A05
C 315	23.4	46.8	1358	8	CC226944	CH261-123	C 388	23.2	46.4	1351	6	CE307000	t1gr-988-
C 316	23.2	46.4	229	2	BB300980	BB300980	C 389	23.2	46.4	1351	6	CD004350	VVA013E02

C 390	23	46.0	573	7	CK179036	EST768356	CK179036	C 453	22.8	45.6	205	9	CE653978	tigr-g889-
C 391	23	46.0	576	9	BR836710	GR0AAA71A	BR836710	464	22.8	45.6	211	1	AA456852	aa38h06.8
C 392	23	46.0	588	8	BR745810	GZ72b007.b	BR745810	C 465	22.8	45.6	212	1	BE496595	NXCI_021
C 393	23	46.0	600	5	BP762554	BP762554	BP762554	466	22.8	45.6	218	1	AV065059	AV065059
C 394	23	46.0	602	5	BP239077	BP239077	BP239077	467	22.8	45.6	250	1	BY354979	BY354979
C 395	23	46.0	605	5	BP762559	BP762559	BP762559	468	22.8	45.6	267	2	BB552157	BB552157
C 396	23	46.0	610	8	BE653267	BOMKN89TR	BE653267	469	22.8	45.6	291	2	BB246033	BB246033
C 397	23	46.0	616	9	CE202027	tigr-g88-	CE202027	C 470	22.8	45.6	307	9	CR794990	GR0AA12C
C 398	23	46.0	625	9	CA2744782	SCMBBb012	CA2744782	471	22.8	45.6	321	2	AW415988	50636 MAR
C 399	23	46.0	637	6	CA235446	SCCCFLA09	CA235446	472	22.8	45.6	326	7	CF367618	842380 MA
C 400	23	46.0	641	9	CR817913	GR0AAA43C	CR817913	C 473	22.8	45.6	327	7	CF368954	853716 MA
C 401	23	46.0	649	9	CC834933	ZMMBBB018	CC834933	474	22.8	45.6	328	9	CNS00690	AG286757
C 402	23	46.0	658	8	BZ522858	BOKAB12TF	BZ522858	475	22.8	45.6	336	9	AG258757	Lotus cor
C 403	23	46.0	659	8	AQ398703	mgx00010D	AQ398703	476	22.8	45.6	340	5	BY488207	BY488207
C 404	23	46.0	696	8	BQ204046	oed14a07	BQ204046	477	22.8	45.6	349	2	BF901436	IL2-WT017
C 405	23	46.0	703	9	AG055269	Pan trogl	AG055269	478	22.8	45.6	357	2	BB725335	BB725335
C 406	23	46.0	708	4	BM437052	VUA013E02	BM437052	479	22.8	45.6	374	4	BI977228	1031048B0
C 407	23	46.0	711	9	AG428556	Mus muscu	AG428556	480	22.8	45.6	397	8	AQ853805	AQ853805
C 408	23	46.0	726	8	BZ037555	oe07c07	BZ037555	C 481	22.8	45.6	398	2	AW001344	AW001344
C 409	23	46.0	732	9	CR831888	GR0AAA64D	CR831888	482	22.8	45.6	405	7	CO115695	GR_EB017
C 410	23	46.0	747	8	BZ059110	llg04905	BZ059110	483	22.8	45.6	413	6	BY532998	BY532998
C 411	23	46.0	753	8	BZ040472	lka38e05	BZ040472	484	22.8	45.6	441	4	BI342114	369789 MA
C 412	23	46.0	758	8	BZ059111	llg04905	BZ059111	C 485	22.8	45.6	441	5	BY158736	BY158736
C 413	23	46.0	761	8	BE650974	BOMOS75TF	BE650974	486	22.8	45.6	463	4	BJ186749	BJ186749
C 414	23	46.0	765	9	CL746224	OR_BBa008	CL746224	487	22.8	45.6	466	6	BY590485	BY590485
C 415	23	46.0	775	8	BZ725679	OGFAJ78TC	BZ725679	488	22.8	45.6	474	4	BH830612	BACPP31-D
C 416	23	46.0	792	4	B1412931	602987380	B1412931	489	22.8	45.6	480	2	BB771345	BB771345
C 417	23	46.0	792	8	BZ056119	lle71f09	BZ056119	490	22.8	45.6	482	7	CF315205	HD--04-A1
C 418	23	46.0	794	9	AG094100	Pan trogl	AG094100	C 491	22.8	45.6	486	2		

```

LIBRARY
  Vector      : pKS145
  R.Site 1   : SacI
  R.Site 2   : SacI.
  Location/Qualifiers
    1..908
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /clone="PTB-078G01.R"
    /sex="male"
    /cell_type="lymphoblast"
    /clone_lib="PTB Chimpanzee Male BAC Library"

FEATURES
  source
    1..908
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /clone="PTB-078G01.R"
    /sex="male"
    /cell_type="lymphoblast"
    /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      55.2%; Score 27.6; DB 9; Length 908;
  Best Local Similarity 72.0%; Pred. No. 99;
  Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CTTACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
  |||||||
Db 597 CCACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 646

RESULT 2
LOCUS      CO881510
DEFINITION BovGen_09835 normal cattle brain Bos taurus cDNA clone
ACCESSION RZPDp1056B1011Q 5', mRNA sequence.
VERSION    CO881510
KEYWORDS   EST.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus

REFERENCE
  1 (bases 1 to 594)
  Hennig.S., Janitz.M., Herwig.R. and Williams.J.
  Generation, annotation, evolutionary analysis and database
  integration of 14969 cattle EST clusters
  Unpublished (2004)
  Contact: Hennig S
  Laboratory 123, dept.Lehrach
  Max-Planck-Institut fuer Molekulare Genetik
  Ihnestr.63-73, D-14195 Berlin, Germany
  Tel: +49 30 8413 1612
  Fax: +49 30 8413 1380
  Email: hennig@molgen.mpg.de
  The library was characterised by oligonucleotide fingerprinting
  (ONFP) to reduce sequencing redundancy. According to the ONFP
  procedure, clones that display the same hybridisation matrix with a
  battery of 200 8mer oligonucleotides are grouped into clusters. One
  clone per ONFP cluster was selected for sequencing. cDNA clones and
  filters are distributed via Deutsches Ressourcenzentrum fuer
  Genomforschung GmbH (http://www.rzpd.de).

PCR PRIMERS
  FORWARD: 5' CCCAGGCTTACATTTATGTTCCGCTCG 3' (M13RSP) 5'-seq
  BACKWARD: 5' CTTATACCCAGCTGGGAAGGGGATGTG 3' (M13RSP) 3'-seq
  Seq primer: 5'-CCGTCGGAAATTCGGGGT-3' (M13RSP).
  Location/Qualifiers
    1..594
    /organism="Bos taurus"
    /mol_type="mRNA"
    /db_xref="taxon:9913"
    /clone="RZPDp1056B1011Q"
    /sex="female"
    /issue_type="brain tissue"
    /dev_stage="adult brain"
    /clone_lib="normal cattle brain"
    /note="Organ: brain; Vector: pSport1; Site:1: NotI;
    Site:2: SalI; Random primed and directionally cloned in
    pSport1 vector using NotI"

FEATURES
  source
    1..594
    /organism="Bos taurus"
    /mol_type="mRNA"
    /db_xref="taxon:9913"
    /clone="RZPDp1056B1011Q"
    /sex="female"
    /issue_type="brain tissue"
    /dev_stage="adult brain"
    /clone_lib="normal cattle brain"
    /note="Organ: brain; Vector: pSport1; Site:1: NotI;
    Site:2: SalI; Random primed and directionally cloned in
    pSport1 vector using NotI"

LIBRARY
  Vector      : pKS145
  R.Site 1   : SacI
  R.Site 2   : SacI.
  Location/Qualifiers
    1..908
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /clone="PTB-078G01.R"
    /sex="male"
    /cell_type="lymphoblast"
    /clone_lib="PTB Chimpanzee Male BAC Library"

FEATURES
  source
    1..908
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /clone="PTB-078G01.R"
    /sex="male"
    /cell_type="lymphoblast"
    /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      54.0%; Score 27; DB 7; Length 594;
  Best Local Similarity 76.7%; Pred. No. 1.5e+02;
  Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 CCACCCCTGGATCTTGGCCCGCCCTTGAATTCCTCAACCTCC 47
  |||||||
Db 280 CACTCCTGGACCTCTGGCCCGCCCTTGGAGACCCACCC 322

RESULT 3
LOCUS      FR0006159
DEFINITION F.rubripes GSS sequence, clone 107H21dA2, genomic survey sequence.
ACCESSION Z89969
VERSION    Z89969.1 GI:1867183
KEYWORDS   GSS; genome survey sequence.
SOURCE     Takifugu rubripes (Fugu rubripes)
ORGANISM   Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.

REFERENCE
  1
  Elgar,G., Clark,M.S., Meek,S., Smith,S., Warner,S., Edwards,Y.J.,
  Bouchireb,N., Cottage,A., Yeo,G.S., Umrانيا,Y., Williams,G. and
  Brenner,S.
  Generation and analysis of 25 Mb of genomic DNA from the pufferfish
  Fugu rubripes by sequence scanning
  Genome Res. 9 (10), 960-971 (1999)
  99455097
  10523524
  2 (bases 1 to 619)
  Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
  Williams,G. and Brenner,S.
  Direct Submission
  Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
  Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
  Vector: pBluescript II KS
  V-type: phagemid
  PRIMER: KS
  DESCR:
  One pass dye-terminator sequencing of cosmid cloned genomic
  sequence.
  Location/Qualifiers
    1..619
    /organism="Takifugu rubripes"
    /mol_type="genomic DNA"
    /db_xref="taxon:31033"
    /clone="107H21dA2"
    /clone_lib="cosmid 107H21"

FEATURES
  source
    1..619
    /organism="Takifugu rubripes"
    /mol_type="genomic DNA"
    /db_xref="taxon:31033"
    /clone="107H21dA2"
    /clone_lib="cosmid 107H21"

ORIGIN
  Query Match      53.2%; Score 26.6; DB 9; Length 619;
  Best Local Similarity 71.4%; Pred. No. 2.1e+02;
  Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
  |||||||
Db 345 CTGCCGACACCTGAGACTTGGCCCGCCCTTGGAGACCCGACGCTCCGC 393

RESULT 4
LOCUS      CL503275/c
DEFINITION SAIL 721_D09.v2 SAIL Collection Arabidopsis thaliana genomic clone
ACCESSION SAIL_721_D09.v2, genomic survey sequence.
VERSION    CL503275
KEYWORDS   CL503275.1 GI:46000595
GSS.

```


Query Match	52.4%	Score 26.2;	DB 6;	Length 960;
Best Local Similarity	72.3%	Pred. No. 2.9e+02;		
Matches 34;	Conservative	0;	Mismatches 13;	Indels 0;
			Gaps	0;

VERSION
KEYWORDS

CK838096.1 GI:45071978
EST.

SOURCE
ORGANISM Bos taurus (cow)
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

REFERENCE
AUTHORS 1 (bases 1 to 365)
Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassell,C.P. and
Matukumalli,L.K.

TITLE
Construction and Analysis of a cDNA Library Generated From
Intestinal Muscle and Epithelial Tissues of Holstein Cattle

JOURNAL
COMMENT Unpublished (2004)
Contact: Richard G. Baumann
Bovine Functional Genomics Lab
ANRI

FEATURES
source
BLDG 162: BARC-EAST, Beltsville, MD 20705, USA
Tel: 3015048604
Fax: 3015048744
Email: rbaumann@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt ' ' -trim fasta. Vector identified
by cross match using options -mismatch 12 -minscore 18
Plate: 1 row: C column: 01
Seq primer: CCTATTAGTGACACTATAGAAC
High quality sequence stop: 365.

ORIGIN
source
1..365
/organism="Bos taurus"
/mol_type="mRNA"
/strains="Holstein"
/db_xref="taxon:9913"
/clone="8BOV 1C01"
/sex="Female"
/tissue_type="Epithelial, Muscle"
/dev_stage="Lactating, Neonatal"
/lab_host="DH10B TonA"
/clone_lib="BARC 8BOV"
/note="Organ: Intestine; Vector: pCMVSPORT6.1; Site 1:
NotI; Site 2: EcoRI; Normalized cow cDNA intestinal
library in pCMVSPORT6.1, constructed from equimolar mRNA
pools derived from 5 sources, 4 lactating intestinal, 1
neonatal intestinal 4/5 lactating, Proximal Duodenum,
Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal
Duodenum, Jejunum, Distal Ileum"

ORIGIN
Query Match 52.0%; Score 26; DB 7; Length 365;
Best Local Similarity 70.0%; Pred. No. 3.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCTGCGCCCTTGAATTCACCTCCAC 50
Db 116 CCCAGCGCTGAGGATCGGGCCAGGACCTGAAATCCAGACCTCAAC 165

RESULT 10
BH639368
LOCUS BH639368 368 bp DNA linear GSS 14-FEB-2002
DEFINITION 100802BH07.2EL_xl 1008 - RescueMu Grid I Zea mays genomic, genomic
survey sequence.

ACCESSION
BH639368
VERSION BH639368.1 GI:18664488
KEYWORDS GSS.

SOURCE
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS 1 (bases 1 to 368)
Walbot,V.

TITLE
Maize genomic sequences found using engineered RescueMu transposon
Contact: Walbot V
Department of Biological Sciences

ORIGIN
Query Match 52.0%; Score 26; DB 8; Length 368;
Best Local Similarity 76.2%; Pred. No. 3.1e+02;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 ACCCACCTCGATCCTTGCCTGCGCCCTTGAATTCACCAACC 45
Db 69 ACCCACCTCGATCCTTGCCTGCGCCCTGAGTAGATTCACGAGCCC 110

RESULT 11
BY605779/c
LOCUS BY605779 386 bp mRNA linear EST 15-DEC-2002
DEFINITION BY605779 RIKEN full-length enriched, visual cortex Mus musculus
cDNA clone K230044B21 3', mRNA sequence.

ACCESSION
BY605779
VERSION BY605779.1 GI:26940961
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 386)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1008028 row: 6
Class: transposon-tagged.

FEATURES
source
Location/Qualifiers
1..368
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1008 - RescueMu Grid I"
/note="Organ: leaf; Vector: RescueMu (engineered from
pBluescript backbone); Site 1: BamHI; Site 2: BglII;
RescueMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescueMu, go to the web
site www.zmldb.iastate.edu and follow the links for
'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."

ORIGIN
Query Match 52.0%; Score 26; DB 8; Length 368;
Best Local Similarity 76.2%; Pred. No. 3.1e+02;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 ACCCACCTCGATCCTTGCCTGCGCCCTTGAATTCACCAACC 45
Db 69 ACCCACCTCGATCCTTGCCTGCGCCCTGAGTAGATTCACGAGCCC 110

RESULT 11
BY605779/c
LOCUS BY605779 386 bp mRNA linear EST 15-DEC-2002
DEFINITION BY605779 RIKEN full-length enriched, visual cortex Mus musculus
cDNA clone K230044B21 3', mRNA sequence.

ACCESSION
BY605779
VERSION BY605779.1 GI:26940961
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 386)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

JOURNAL
COMMENT

Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:31295370.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
9773.f

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0DI053BE08NP1&c=9773.f.

FEATURES

source

Location/Qualifiers

1..899

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="CS0DI053YI16"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 52.0%; Score 26; DB 1; Length 699;
Best Local Similarity 71.7%; Pred. No. 3.3e+02;
Matches 33; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 5 CCACCTGGATCTGCGCCCGCCCTTGAATTCACACCTCCAC 50

Db 27 CACAGCTGAACCTTVCACCCCGCTGTCTCCACCCCGCCAC 72

RESULT 14
A0253286

LOCUS HS_2046_B1_D02_MR_CIT_Approved_Human_Genomic_Sperm_Library_D_Homo
DEFINITION sapiens genomic clone Plate=2046 Col=3 Row=H, genomic survey
sequence.

ACCESSION A0253286

VERSION A0253286.1

KEYWORDS GI:3724640

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE 1 (bases 1 to 726)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

PUBMED 10449764

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2046 row: H column: 3

Class: BAC ends

High quality sequence stop: 726.

Location/Qualifiers

1..726

/organism="Homo sapiens"

FEATURES

source

Query Match 52.0%; Score 26; DB 7; Length 739;

Best Local Similarity 70.0%; Pred. No. 3.3e+02;

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone_lib="Plate=2046 Col=3 Row=H"

/sex="male"

/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBLOBAC11; BAC Clones in
E-Coli DH10B"

ORIGIN

Query Match 52.0%; Score 26; DB 8; Length 726;
Best Local Similarity 70.0%; Pred. No. 3.3e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCTGTCGCCCGCCCTTGAATTCACACCTCCAC 50

Db 296 CTCCCCCCCCCTCTCTCCCGCTCCCGCTTGCCTCCCCCTCCCTCCC 345

RESULT 15

CK833690

LOCUS

DEFINITION

CK833690

ACCESSION

CK833690

VERSION

CK833690.1

KEYWORDS

EST.

SOURCE

Bos taurus (cow)

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

REFERENCE 1 (bases 1 to 739)

AUTHORS

Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassell,C.P. and

Matukumalli,L.K.

TITLE

Construction and Analysis of a cDNA Library Generated From

Intestinal Muscle and Epithelial Tissues of Holstein Cattle

Unpublished (2004)

JOURNAL

COMMENT

Contact: Richard G. Baumann

Bovine Functional Genomics Lab

ANRI

BLDG 162: BARC-EAST, Beltsville, MD 20705, USA

Tel: 3015048604

Fax: 3015048744

Email: rbaumann@nri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

0.000925 using options -trim_alt -trim_fasta. Vector identified

by cross match using options -minmatch 12 -minscore 12

Plate: 3 row: M column: 09

Seq primer: CCTATTAGTGACACTATAGAAC

High quality sequence stop: 739.

FEATURES

Location/Qualifiers

1..739

/organism="Bos taurus"

/mol_type="mRNA"

/strain="Holstein"

/db_xref="taxon:9913"

/clone_lib="Bos 8BOV"

/sex="female"

/tissue_type="Epithelial, Muscle"

/dev_stage="Lactating, Neonatal"

/lab_host="DH10B TonA"

/clone_lib="Bos 8BOV"

/note="Organ: Intestine; Vector: pCMVSPORT6.1; Site_1:

NotI; Site_2: EcoRI; Normalized cow cDNA intestinal

library in pCMVSPORT6.1, constructed from equimolar mRNA

pools derived from 5 sources, 4 lactating intestinal, 1

neonatal intestinal 4/5 Lactating, Proximal Duodenum,

Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal

Duodenum, Jejunum, Distal Ileum"

Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 CCTACCCACCTGGATCCTTGGCCCCGCCCTTGAATTTCCCAACCTCCAC 50
 |||||
 Db 157 CCCAGCCGCTGAGGATCGGSCCCAGGACCTGAAATTTCCCGAGCTCAAC 206
 |||||

RESULT 16
 ALS522831 877 bp mRNA linear EST 24-MAR-2004
 LOCUS
 DEFINITION ALS522831 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
 cDNA clone CS0DB009YP11 3-PRIME, mRNA sequence.
 ACCESSION ALS522831
 VERSION
 KEYWORDS
 SOURCE ALS522831.2 GI:31041095
 EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12786324.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 9773.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0DB009YCH06NP1&c=9773.f.
 Location/Qualifiers
 1. .877
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DB009YP11"
 /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
 Query Match 52.0% Score 26; DB 1; Length 877;
 Best Local Similarity 69.6% Pred.No. 3.3e+02;
 Matches 33; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 5 CCCACCTGGATCCTTGGCCCCGCCCTTGAATTTCCCAACCTCCAC 50
 |||||
 Db 30 CACAGCTTGGSCCTTSSCCACCCCTTGTCCTCCCAACCCCAAC 75
 |||||

RESULT 17
 AG331887/c
 LOCUS
 DEFINITION AG331887 Mus musculus molossinus DNA, clone:MSG01-121005.T7, genomic survey
 sequence.
 ACCESSION AG331887
 VERSION
 KEYWORDS
 SOURCE AG331887.1 GI:47905197
 GSS.
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kotima, T. and Sakaki, Y.

BAC end Sequences of Library MSMg01

Unpublished
2 (bases 1 to 966)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submision
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Teukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS

Sequencing : T7
Library
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI

Location/Qualifiers
1..966
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-121005.T7"
/sex="male"
/tissue_types="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 52.0%; Score 26; DB 9; Length 966;
Best Local Similarity 70.0%; Pred.No. 3.4e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCCTCGATCTTGTCCCGCCCGCCCTTGAAATTCACCAACCTCCAC 50
Db 405 CCTACCCCGCCCTTCCCGCTACCCCGCCCGCCCTCCTCCTCCACCCCTCCTC 356

RESULT 18

BM470168

LOCUS
DEFINITION
AGENCOURT_6463404 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5538947
5', mRNA sequence.

ACCESSION
VERSION
BM470168
BM470168.1 GI:18519210

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 1057)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS
TITLE
JOURNAL

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL12232 row: k column: 12
High quality sequence stop: 563.
Location/Qualifiers
1..1057

FEATURES
source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5538947"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: Not1; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN
Query Match      52.0%; Score 26; DB 4; Length 1057;
Best Local Similarity 76.2%; Pred. No. 3.4e+02;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 CCACCCCTGGATCCTTGGCCCGCCCTTGAATTCACCAACCCCT 46
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 889 CCACCGTGCTCTCTCCCGCCCGCCCTTGTGTCACCAACCCCT 930

RESULT 19
LOCUS      CF094505      283 bp      mRNA      linear      EST 22-JUL-2003
DEFINITION Helianthus argophyllus cDNA clone QHN18C01, mRNA sequence.
ACCESSION  CF094505
VERSION     CF094505.1 GI:33133572
KEYWORDS   EST.
SOURCE     Helianthus argophyllus
ORGANISM   Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE  1 (bases 1 to 283)
AUTHORS   Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
            Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
            Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
            Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
JOURNAL   Unpublished (2002)
COMMENT   Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmudson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig1015, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHN18 row: C column: 01.

FEATURES
            source
            Location/Qualifiers
                1..283
                /organism="Helianthus argophyllus"
                /mol_type="mRNA"
                /db_xref="taxon:73275"
                /clone="QHN18C01"
                /lab_host="E.coli"
                /clone_lib="QH N sunflower H. argophyllus (drought stress)"
                /notes="Vector: pGEM-T; The library was constructed from
                three different sources (seedling, root and leaf) of RNA
                from a single genotype. cDNAs were pooled and cloned into
                a high-copy vector pGEM-T. Details of library construction
                can be obtained at http://cgpdb.ucdavis.edu/"

ORIGIN
Query Match      51.6%; Score 25.8; DB 7; Length 283;
Best Local Similarity 73.3%; Pred. No. 3.5e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5538947"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: Not1; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN
Query Match      52.0%; Score 26; DB 4; Length 1057;
Best Local Similarity 76.2%; Pred. No. 3.4e+02;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 CCACCCCTGGATCCTTGGCCCGCCCTTGAATTCACCAACCCCT 46
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 889 CCACCGTGCTCTCTCCCGCCCGCCCTTGTGTCACCAACCCCT 930

RESULT 20
LOCUS      BQ943145      933 bp      mRNA      linear      EST 21-AUG-2002
DEFINITION AGENCOURT 8795449 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6429868
            5', mRNA sequence.
ACCESSION  BQ943145
VERSION     BQ943145.1 GI:22358623
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 933)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            Tissue Procurement: DCTD/DRP/Gazdar
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2618 row: m column: 05
            High quality sequence start: 38
            High quality sequence stop: 600.

FEATURES
            source
            Location/Qualifiers
                1..933
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6429868"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_18"
                /notes="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(G). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH_MGC Library."

ORIGIN
Query Match      51.6%; Score 25.8; DB 5; Length 933;
Best Local Similarity 81.1%; Pred. No. 3.9e+02;
Matches 30; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 14 GATCCTTGGCCCGCCCTTGAATTCACCAACCCCTCCAC 50
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 876 GATCCTGGACCCGTCCTCCATGATTCACCAACCCCTCCC 912

RESULT 21
LOCUS      CE491137      309 bp      DNA      linear      GSS 28-SEP-2003
DEFINITION tigr-gss-dog-17000365592748 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION  CE491137
VERSION     CE491137.1 GI:36807918
KEYWORDS   GSS.
SOURCE     Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

```

```

REFERENCE
AUTHORS      1 (bases 1 to 309)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.

TITLE
JOURNAL      The dog genome: survey sequencing and comparative analysis
MEDLINE      Science 301 (5641), 1898-1903 (2003)
PUBMED      22875432
PUBMED      14512627

COMMENT
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..309
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match      51.2%; Score 25.6; DB 9; Length 309;
Best Local Similarity 70.8%; Pred. No. 4.1e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CTTACCCACCTGGATCTTGGCCGCCCTTGAATTCCTCCACCTCC 48
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 204 CTTAACTTCCCTGGAATCCAGCCGCCCTTGGCTGCTAGGCTCC 157

RESULT 22
CN438735/c
LOCUS      388 bp mRNA linear EST 07-JUL-2004
DEFINITION BE04017A1A05 BE04 Normalized and Subtracted bovine embryonic and
extraembryonic tissue Bos taurus cDNA clone BE04017A1A05 5', mRNA
sequence.
CN438735
ACCESSION CN438735.1 GI:46417999
VERSION EST.
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 388)
Lewin,H.A., Renard,J.P., Yang,X.J., Hernandez,A., Degrelle,S.,
Hue,I., Tian,X.C., Liu,L. and Everts,R.E.
Bovine embryonic ESTs
Unpublished (2004)
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana-Champaign
206 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for embryonic EST sequencing was provided by USDA ARS
Contract No. 58-1275-2-020 to H. A. Lewin. Base Calling/Quality
Scores: PHRED from Washington University Genome Center. Vector
Trimming: Cross_match from Washington University Genome Center
PHRAP suite. Sequences submitted are vector free and at least 200
bp in length with average PHRED score > 20.
PCR primers
FORWARD: CCACTCAGACGCTTGTAAACGAC (M13 NTA-F)
BACKWARD: GTGTGGAACTGTGACGGATAACAA (M13 NTA-R)
Insert Length: 388 Std Error: 0.00

```

```

Plate: BE04017A1 row: A column: 05
Seq primer: TAATACGACTCACTATAGG (T7 PROMOTER)
High quality sequence stop: 388.
Location/Qualifiers
1..388
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="BE04017A1A05"
/tissue_type="embryo (day 36 and day 64) and
extra-embryonic tissue (day 14 to 25) Normalized and twice
subtracted library"
/lab_host="DH108"
/clone_lib="BE04 Normalized and Subtracted bovine
embryonic and extraembryonic tissue"
/note="Vector: pGEMZf1(+); Site 1: EcoRI; Site 2: NotI;
The cDNA library was constructed, normalized and
subtracted by Dr. A. Hernandez, W.M. Keck Center,
University of Illinois Urbana-Champaign and S. Degrelle,
BDR, Institut National de la Recherche Agronomique,
France, as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. (1996), Genome Research 6(9): 791-806. This
clone is from library BE04, a normalized library of
embryonic tissues subtracted with 7800 known placental EST
sequences and 5000 clones from the BE03 library. The
double stranded cDNA was size selected (more than 450 bp),
adapted with EcoRI adaptors at both ends and then
digested with NotI. The cDNA was then directionally cloned
into EcoRI-NotI digested pGEM11zf(+) phagemid vector.
Insert size was between 450-2200 bp. Sequencing was done
from the 5' end of the clone. Tagged oligo-dT was used to
identify the source of the ESTs where possible. (A18)TCGGT
= extraembryonic tissue; (A18)TCGGT = 36-day old fetus;
(A18)TCGGA = 64-day old fetus (organs); (A18)TCGGA =
64-day old fetus (body)"

ORIGIN
Query Match      51.2%; Score 25.6; DB 7; Length 388;
Best Local Similarity 70.8%; Pred. No. 4.2e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCACCTGGATCTTGGCCGCCCTTGAATTCCTCCACCTCCCA 49
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 CCACCTCCCTGGTCTTGGCCACCCCTCTCTCTCACTCCCTCCA 217

RESULT 23
BY356513
LOCUS      444 bp mRNA linear EST 12-DEC-2002
DEFINITION BY356513 RIKEN full-length enriched, ES cells Mus musculus cDNA
clone 2410118L06 3', mRNA sequence.
BY356513
ACCESSION BY356513.1 GI:26586001
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Chordata; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 444)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Oeato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Bacalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chotha,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Flatcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Glessi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagai,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,

```

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES

source

Location/Qualifiers
1..444
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2410118L08"
/cell_type="ES cells"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, ES cells"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',

GAGAGAGAGAGATCCCAAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5',

ORIGIN

Query Match 51.2%; Score 25.6; DB 5; Length 444;
Best Local Similarity 70.8%; Pred. No. 4.3e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 3 TACCACCCCTGGATCTCTTGGCCCGCCCTTGAATTCCTCAACCCCTCCAC 50
Db 128 TCCCACTCCATTCATGCCCCACCCCTTGGACTCTCTATCCTTACAC 175

RESULT 24

BQ953457/c

LOCUS BQ953457

DEFINITION AGENCOURT_8818345 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6390607

5', mRNA sequence.

ACCESSION BQ953457

VERSION BQ953457.1

KEYWORDS GI:22368935

SOURCE EST.

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

NIH-MGC <http://mgc.nci.nih.gov/>.

1 (bases 1 to 1164)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13877 row: i column: 08

High quality sequence start: 136

High quality sequence stop: 432.

Location/Qualifiers

1..1164

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:6390607"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC_129"

/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;

Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.

Primer: Oligo dT. Average insert size 2-2 kb. Constructed

by ResGen, Invitrogen Corp. Note: this is a NIH_MGC

Library."

ORIGIN

Query Match 51.2%; Score 25.6; DB 5; Length 1164;
Best Local Similarity 77.5%; Pred. No. 4.6e+02;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CTTACCCACCTGGATCTTCTCCCGCCCTTGAATTCCTCC 40
Db 40 CTTCCCGCGAGCGGTGCGTCCCGCCCTTCCATTCCTCC 1

RESULT 25

CL051298

LOCUS CL051298

DEFINITION CH216-73K18 Sp6.1 CH216 Xenopus tropicalis genomic clone

CH216-73K18, genomic survey sequence.

ACCESSION CL051298

VERSION CL051298.1

KEYWORDS GI:40507211

SOURCE GSS.

Xenopus tropicalis (western clawed frog)

CL051298 1211 bp DNA linear GSS 31-DEC-2003

```

ORGANISM      Xenopus tropicalis
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
               Xenopodinae; Xenopus; Silurana.
REFERENCE
AUTHORS      Krenitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
               Mardis, E. and Wilson, R.
TITLE        A physical map of the xenopus tropicalis genome
JOURNAL      Unpublished (2003)
COMMENT      Contact: Richard K Wilson
               Genome Sequencing Center
               Washington University School of Medicine
               Email: submissions@watson.wustl.edu
               Insert Length: 175000 Std Error: 0.00
               Seq primer: Sp6 ATTTAGTGACACTATAG
               Class: BAC ends
               High quality sequence start: 11
               High quality sequence stop: 193.
               Location/Qualifiers
                 1..1211
                   /organism="Xenopus tropicalis"
                   /mol_type="genomic DNA"
                   /strain="Nigerian frog"
                   /db_xref="taxon:8364"
                   /clone="CH216-73K18"
                   /sex="male"
                   /cell_line="Stock 248 F7A2, inbred N7"
                   /clone_lib="CH216"
                   /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                   BAC library"
ORIGIN
Query Match      51.2%; Score 25.6; DB 9; Length 1211;
Best Local Similarity 70.8%; Pred. No. 4.7e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1  CCTACCCACCTGGATCTTGGCCCGCCCTTGAAATTCACACCCCTCC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      591  CTTCCATCTCTCCATCCCTCCACCACCCCTTCATCTCCACCCCCC 638

RESULT 26
CE229900/c
LOCUS      tigr-gss-dog-1700033282555 Dog Library Canis familiaris genomic,
DEFINITION      genomic survey sequence.
ACCESSION      CE229900
VERSION        CE229900.1 GI:35385658
KEYWORDS       GSS.
SOURCE         Canis familiaris (dog)
ORGANISM       Canis familiaris
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
AUTHORS      Kirkness, E.F., Barna, V., Halpern, A.L., Levy, S., Remington, K.,
               Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
               Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED       14512627
COMMENT      Contact: Kirkness EF
               The Institute for Genomic Research
               Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
               Rockville, MD 20850, USA
               Tel: 301-838-0200
               Fax: 301-838-0208
               Email: ekirknes@tigr.org
               Class: shotgun.
               Location/Qualifiers
                 1..192
                   /organism="Canis familiaris"
                   /mol_type="genomic DNA"
ORIGIN
Query Match      51.2%; Score 25.6; DB 9; Length 1211;
Best Local Similarity 70.8%; Pred. No. 4.7e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1  CCTACCCACCTGGATCTTGGCCCGCCCTTGAAATTCACACCCCTCC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      591  CTTCCATCTCTCCATCCCTCCACCACCCCTTCATCTCCACCCCCC 638

RESULT 26
CE229900/c
LOCUS      tigr-gss-dog-1700033282555 Dog Library Canis familiaris genomic,
DEFINITION      genomic survey sequence.
ACCESSION      CE229900
VERSION        CE229900.1 GI:35385658
KEYWORDS       GSS.
SOURCE         Canis familiaris (dog)
ORGANISM       Canis familiaris
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
AUTHORS      Kirkness, E.F., Barna, V., Halpern, A.L., Levy, S., Remington, K.,
               Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
               Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED       14512627
COMMENT      Contact: Kirkness EF
               The Institute for Genomic Research
               Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
               Rockville, MD 20850, USA
               Tel: 301-838-0200
               Fax: 301-838-0208
               Email: ekirknes@tigr.org
               Class: shotgun.
               Location/Qualifiers
                 1..192
                   /organism="Canis familiaris"
                   /mol_type="genomic DNA"
ORIGIN
Query Match      50.4%; Score 25.2; DB 9; Length 192;
Best Local Similarity 71.7%; Pred. No. 5.4e+02;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      5  CCCACCTGGATCTTGGCCCGCCCTTGAAATTCACACCCCTCCAC 50
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      158  CCCACCCATATCTTCTTCCACCATGAGCCCCCCCCCTCCAC 113

RESULT 27
CK840615/c
LOCUS      UI-R-AFl-aar-c-02-0-UI-s10 UI-R-AFl Rattus norvegicus cDNA clone
DEFINITION      UI-R-AFl-aar-c-02-0-UI 3', mRNA sequence.
ACCESSION      CK840615
VERSION        CK840615.1 GI:45193600
KEYWORDS       EST.
SOURCE         Rattus norvegicus (Norway rat)
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE
AUTHORS      Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED       8899548
COMMENT      Contact: Soares, MB
               Coordinated Laboratory for Computational Genomics
               University of Iowa
               375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: bento-soares@uiowa.edu
               CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
               CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: Distribution information can be found at
               http://genome.uiowa.edu/distribution/rat.html
               The following repetitive elements were found in this cDNA
               sequence: 390-426, >MST-INTERNAL#LTR/MALR (matched complement)
               Seq primer: M13 FORWARD
               POLVA=Yes.
               Location/Qualifiers
                 1..601
                   /organism="Rattus norvegicus"
                   /mol_type="mRNA"
                   /strain="Sprague-Dawley"
                   /db_xref="taxon:10116"
                   /clone="UI-R-AFl-aar-c-02-0-UI"
                   /dev_stage="adult"
                   /lab_host="DH10B (Life Technologies)"
                   /clone_lib="UI-R-AFl"
                   /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                   polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-AFl
                   library is a normalized library constructed from 15 dpc
                   rat atrioventricular (AV) canal. The tag is a string of 5
                   nucleotides present between the Not I site and the
                   oligo-dr track. The library was constructed as described
                   by Bonaldo, Lennon and Soares, Genome Research 6:
                   791-806, 1996. Tissue provided by Jim Lin, Department of
                   Biology, University of Iowa.
                   TAG_TISSUE=AV canal at 15 dpc
                   TAG_LIB=UI-R-AFl

```

TAG_SEQ=GAAGG

ORIGIN

Query Match 50.4%; Score 25.2; DB 7; Length 601;
 Best Local Similarity 78.9%; Pred. No. 5.9e+02;
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTACCCACCCCTGGATCTTGCCTGCCCGCCCTTGAATTC 38
 |||||
 Db 519 CCTATCGCTGGATCTTGTGCCACACCCCTGAGTTC 482
 |||||

RESULT 28

CE073587/c

LOCUS tigr-gss-dog-17000323611465 Dog Library Canis familiaris genomic,
 DEFINITION genomic survey sequence.

ACCESSION CE073587

VERSION CE073587.1 GI:35137641

KEYWORDS

GSS.

SOURCE

Canis familiaris (dog)

ORGANISM

Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 727)
 Kirkses E.F., Bafna V., Halpern A.L., Levy S., Remington K.,
 Rusch D.B., Delcher A.L., Pop M., Wang W., Fraser C.M. and
 Venter J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

MEDLINE 22875432

PUBMED 14512627

COMMENT Contact: Kirkness EP

The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

FEATURES

source

Location/Qualifiers

1..727
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from
 peripheral blood"

ORIGIN

Query Match 50.4%; Score 25.2; DB 9; Length 727;
 Best Local Similarity 71.7%; Pred. No. 6e+02;
 Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 5 CCCACCTGGATCTTGCCTGCCCGCCCTTGAATTCACACCTCCAC 50
 |||||
 Db 133 CCCACCTGGAGGTGGGCCACCCCGAGGCTTCCCTCCCTCCAC 88
 |||||

RESULT 29

CK446005

LOCUS pncs915aG04.SP6 Aspergillus nidulans negative subtraction
 DEFINITION hybridization polysaccharide related cDNA plasmid library
 Emericella nidulans cDNA clone pncs915aG04 5', similar to
 gi|1173091|sp|P41391|RNA1 SCHPO RAN GTPASE ACTIVATING PROTEIN 1
 (RNA1 PROTEIN), mRNA sequence.

ACCESSION CK446005

VERSION CK446005.1 GI:40883488

KEYWORDS

EST.

SOURCE

Emericella nidulans (anamorph: Aspergillus nidulans)

ORGANISM

Emericella nidulans

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; Emericella.
 1 (bases 1 to 758)

Ray, A., Macwana, S., Ayoubi, P., Hall, J. T., Prade, R. and Mort, A. J.
 Negative subtraction hybridization: An efficient method to isolate
 large numbers of condition-specific cDNAs
 BMC Genomics 5, 22 (2004)

Contact: Patricia Ayoubi

Department of Biochemistry and Molecular Biology

Oklahoma State University

248 Noble Research Center, Stillwater OK, 74078

Tel: 405-744-6209

Fax: 405-744-7799

Email: ayoubi@okstate.edu

PCR Primers

FORWARD: SP6

BACKWARD: T7

Insert Length: 1 Std Error: 0.00

Seq primer: SP6.

FEATURES

source

Location/Qualifiers

1..758

/organism="Emericella nidulans"

/mol_type="mRNA"

/strain="FGSC C26"

/db_xref="taxon:162425"

/clone="pncs915aG04"

/tissue_type="vegetative mycelia"

/lab_host="E. coli"

/clone_lib="Aspergillus nidulans negative subtraction
 hybridization polysaccharide related cDNA plasmid library"
 /note="Vector: pCMVSPORT6.0; Site 1: EcoRI; Site 2:
 HindIII; 5' end of cDNA cloned near EcoRI site of
 pCMVSPORT6.0 and 3' end cloned near HindIII site of
 pCMVSPORT6.0. Average length of insert is 1.49 kb"

ORIGIN

Query Match 50.4%; Score 25.2; DB 7; Length 758;
 Best Local Similarity 71.7%; Pred. No. 6.1e+02;
 Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 5 CCCACCTGGATCTTGCCTGCCCGCCCTTGAATTCACACCTCCAC 50

Db 353 CCTCACCTTCTCTCTTAATGCGCTTGAATTCACACCTCCAC 398

RESULT 30

AG313344

LOCUS

DEFINITION

AG313344 855 bp DNA linear GSS 02-JUN-2004
 Mus musculus molossinus DNA, clone: MSMG01-096L19.T7, genomic survey
 sequence.

ACCESSION

AG313344

VERSION

AG313344.1 GI:47886298

KEYWORDS

GSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

BAC end Sequences of Library MSMG01

2 (bases 1 to 855)

Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@gsr.riken.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the mouse BAC library MSMG01. For BAC
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Query Match 50.0%; Score 25; DB 2; Length 556;
Best Local Similarity 69.4%; Pred. No. 6.9e+02;

JOURNAL
COMMENT

Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.
Seq primer: AGCGATAACAATTTCCACACGGA.
Location/Qualifiers
1. .617
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="IRT22A16 C A08"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCWA lrt"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."

ORIGIN

Query Match 50.0%; Score 25; DB 6; Length 617;
Best Local Similarity 75.6%; Pred. NO. 6.e+02;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 CACCTGGATCTTGGCCGCGCCCTTGAATTCCTCCACCCCTC 47
|||||
DB 259 CACTCGGGTCTTGTGTGCGCTCCCTGCATCCCTAACCTC 219
|||||

RESULT 37
LOCUS CK846307/c

DEFINITION 968792 MARC 4BOV Bos taurus cDNA 3', mRNA linear EST 05-MAR-2004

ACCESSION CK846307

VERSION CK846307.1 GI:45206692

KEYWORDS

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

REFERENCE

AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perlea,G., Holt,I., Karamycheva,S., Liang,P., Quackenbush,J. and Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

PUBMED 11282978

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: 64 row: M column: 1
Seq primer: GTAATCAGCTACTACTAGGG.
Location/Qualifiers
1. .652
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"

FEATURES

source

Query Match 50.0%; Score 25; DB 9; Length 678;
Best Local Similarity 69.4%; Pred. NO. 7e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 CTACCACCTGGATCTTGGCCGCGCCCTTGAATTCCTCCACCCCTCCAC 50
|||||
DB 470 CTGGCACCTCCACCCGCGCCGAGCCCTCAAAATACCCACCCCTCAAC 422
|||||

RESULT 38
LOCUS AG568473

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-490C17.TJ, genomic survey sequence.

ACCESSION AG568473

VERSION AG568473.1 GI:48329171

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

REFERENCE

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE BAC end Sequences of Library MSMg01

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 678)

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel.81-45-503-9111, Fax:81-45-503-9170)

COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.site 1 : EcoRI.

R.site 2 : EcoRI.

Location/Qualifiers
1. .678
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-490C17.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

FEATURES

source

Query Match 50.0%; Score 25; DB 9; Length 678;
Best Local Similarity 69.4%; Pred. NO. 7e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCACCTGGATCTTGGCCGCGCCCTTGAATTCCTCCACCCCTCCA 49
|||||
DB 545 CTTTCCCCCTCCACCCGCGCCGCTTTAATCCCCCCCCCCCCCA 593
|||||

JOURNAL
COMMENT

Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.
Seq primer: AGCGATAACAATTTCCACACGGA.
Location/Qualifiers
1. .617
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="IRT22A16 C A08"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCWA lrt"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."

ORIGIN

Query Match 50.0%; Score 25; DB 7; Length 652;
Best Local Similarity 69.4%; Pred. NO. 7e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 CTACCACCTGGATCTTGGCCGCGCCCTTGAATTCCTCCACCCCTCCAC 50
|||||
DB 470 CTGGCACCTCCACCCGCGCCGAGCCCTCAAAATACCCACCCCTCAAC 422
|||||

RESULT 38
LOCUS AG568473

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-490C17.TJ, genomic survey sequence.

ACCESSION AG568473

VERSION AG568473.1 GI:48329171

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

REFERENCE

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE BAC end Sequences of Library MSMg01

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 678)

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel.81-45-503-9111, Fax:81-45-503-9170)

COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.site 1 : EcoRI.

R.site 2 : EcoRI.

Location/Qualifiers
1. .678
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-490C17.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

FEATURES

source

Query Match 50.0%; Score 25; DB 9; Length 678;
Best Local Similarity 69.4%; Pred. NO. 7e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCACCTGGATCTTGGCCGCGCCCTTGAATTCCTCCACCCCTCCA 49
|||||
DB 545 CTTTCCCCCTCCACCCGCGCCGCTTTAATCCCCCCCCCCCCCA 593
|||||

```

RESULT 39
BX297790/c
LOCUS
DEFINITION BX297790 tcay Oncorhynchus mykiss cDNA clone tcay0003b.a.24 5prim, linear EST 10-MAY-2004
ACCESSION BX297790
VERSION
KEYWORDS
SOURCE
ORGANISM
Oncorhynchus mykiss (rainbow trout)

REFERENCE
AUTHORS Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE Construction and primary characterization of normalized cDNA
JOURNAL libraries in rainbow trout, Oncorhynchus mykiss
COMMENT Unpublished (2003)
On Apr 7, 2003 this sequence version replaced gi:29578435.
Contact: Guiguen Y
INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at signenausupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0003 row: a column: 24
Seq primer: M13R.

FEATURES
source
1..788
Location/Qualifiers
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcay0003b.a.24"
/tissue_type="adipose tissue, blood, brain,
differentiating gonads, gills, interrenal, intestine,
kidney, liver, muscle, ovary, pituitary, testis"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="tcay"
/note="Vector: pT73D-pac; Rainbow trout multi-tissues -
normalized + 1 subtraction (tcay); Clone distribution :
AGENAE Resource centre, Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouv-en-Josas cedex, France, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

ORIGIN
Query Match 50.0%; Score 25; DB 5; Length 788;
Best Local Similarity 75.6%; Pred. No. 7.1e+02;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 CACCTGTGATCCTTCCCGCCCTTGTAATTCCTCAACCTC 47
|||||
DB 617 CACTCGGGGTCTTTCGGCTCCCTGTCATCCCTCAACCTC 577
|||||

RESULT 40
BZ153219
LOCUS
DEFINITION BZ153219 TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
ACCESSION BZ153219
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)

REFERENCE
AUTHORS Rattus norvegicus
TITLE Rattus norvegicus
JOURNAL Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT GSS.

QY 7 CACCTGTGATCCTTCCCGCCCTTGTAATTCCTCAACCTC 47
|||||
DB 617 CACTCGGGGTCTTTCGGCTCCCTGTCATCCCTCAACCTC 577
|||||

RESULT 41
BZ036939
LOCUS
DEFINITION BZ036939 MP Mesembryanthemum crystallinum cDNA 5' similar to
ACCESSION BZ036939
VERSION BZ036939.1 GI:8331955
KEYWORDS
SOURCE
ORGANISM
Mesembryanthemum crystallinum (common iceplant)

REFERENCE
AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
Ferrea,H., Kawasaki,S., McColough,A., Michalowski,C.B.,
Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE Functional Genomics of Plant Stress Tolerance
JOURNAL Unpublished (2000)
COMMENT Contact: Michalowski,C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel.:520-621-7982
Fax: 520-621-1697

```

```

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 815)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other GSSs: CH230-360C9.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
plate: http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html
Plate: 360 row: C column: 9
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..815
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-360C9"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match 50.0%; Score 25; DB 8; Length 815;
Best Local Similarity 69.4%; Pred. No. 7.1e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCACCTCGATCTTCCCGCCCTTGTAATTCCTCAACCTCCA 49
|||||
DB 685 CCCATACACTGCTGACCTTCCCTTCCCTTGGAGTCCCAACCTCCA 733
|||||

RESULT 41
BZ036939
LOCUS
DEFINITION BZ036939 MP Mesembryanthemum crystallinum cDNA 5' similar to
ACCESSION BZ036939
VERSION BZ036939.1 GI:8331955
KEYWORDS
SOURCE
ORGANISM
Mesembryanthemum crystallinum (common iceplant)

REFERENCE
AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
Ferrea,H., Kawasaki,S., McColough,A., Michalowski,C.B.,
Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE Functional Genomics of Plant Stress Tolerance
JOURNAL Unpublished (2000)
COMMENT Contact: Michalowski,C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel.:520-621-7982
Fax: 520-621-1697

```



```
/organism="Mus musculus"
/mol_type="mRNA"
/strains="129, C57BL/6J, FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5291035"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
```

ORIGIN

```
Query Match          49.6%; Score 24.8; DB 4; Length 724;
Best Local Similarity 72.7%; Pred. No. 8.2e+02;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 CCACCTGGATCCTTGCCTCCCGCCCTTGAATTCACACCTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 CCTGCCTGGACCGCGCCCTCCCGTGAACCTCCCTTC 175
```

RESULT 49

```
CD300548
LOCUS              CD300548      876 bp      mRNA      linear      EST 27-MAY-2003
DEFINITION        AGENCOURT 14255375 NIH MGC 180 Homo sapiens cDNA clone
IMAGE:30389460 5', mRNA sequence.
ACCESSION         CD300548
VERSION           CD300548.1 GI:31080343
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 876)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Daniela S. Gerhard, Ph.D.
                  Office of Cancer Genomics
                  National Cancer Institute / NIH
                  Bldg. 31 Rm10A07 Bethesda, MD 20892
                  Email: c9apbs-remail.nih.gov
                  cDNA Library Preparation: Invitrogen Corp
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: NDAM462 row: p column: 13
                  High quality sequence stop: 614.
```

FEATURES

```
source
Location/Qualifiers
1..876
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30389460"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."
```

ORIGIN

```
Query Match          49.6%; Score 24.8; DB 6; Length 876;
Best Local Similarity 72.7%; Pred. No. 8.4e+02;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 ACCACCTGGATCCTTGCCTCCCGCCCTTGAATTCACACCTCC 47
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 805 ACCCCCTTGAATTTCCCGCCCATTTGTTTCCAGCCCC 848
```

Search completed: April 25, 2005, 15:04:02
Job time : 4328.59 secs

```
Best Local Similarity 72.7%; Pred. No. 8.3e+02;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 ACCACCTGGATCCTTGCCTCCCGCCCTTGAATTCACACCTCC 47
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 782 ACCCCCTTGAATTTCCCGCCCATTTGTTTCCAGCCCC 825
```

RESULT 50

```
CD557122
LOCUS              CD557122      947 bp      mRNA      linear      EST 11-JUN-2003
DEFINITION        AGENCOURT 14413435 NIH MGC 180 Homo sapiens cDNA clone
IMAGE:30390355 5', mRNA sequence.
ACCESSION         CD557122
VERSION           CD557122.1 GI:31583190
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 947)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Daniela S. Gerhard, Ph.D.
                  Office of Cancer Genomics
                  National Cancer Institute / NIH
                  Bldg. 31 Rm10A07 Bethesda, MD 20892
                  Email: c9apbs-remail.nih.gov
                  cDNA Library Preparation: Invitrogen Corp
                  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: NDAM465 row: e column: 20
                  High quality sequence start: 9
                  High quality sequence stop: 495.
```

FEATURES

```
source
Location/Qualifiers
1..947
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390355"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."
```

ORIGIN

```
Query Match          49.6%; Score 24.8; DB 6; Length 947;
Best Local Similarity 72.7%; Pred. No. 8.4e+02;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 ACCACCTGGATCCTTGCCTCCCGCCCTTGAATTCACACCTCC 47
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 805 ACCCCCTTGAATTTCCCGCCCATTTGTTTCCAGCCCC 848
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:14:29 ; Search time 150 Seconds
(without alignments)
545.426 Million cell updates/sec

Title: US-10-010-476-13
Perfect score: 50
Sequence: 1 CTTACCCACCTGGATCTT.....TTGAATTCACACCTTCAC 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A COMB.seq:
2: /cgn2_6/prodata/1/ina/5B COMB.seq:
3: /cgn2_6/prodata/1/ina/6A COMB.seq:
4: /cgn2_6/prodata/1/ina/6B COMB.seq:
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq:
6: /cgn2_6/prodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.6	53.2	67002	4	US-09-949-016-15803
2	24.4	48.8	101951	4	US-09-949-016-15648
3	23.6	47.2	45275	4	US-09-949-016-12533
4	23.4	46.8	60572	4	US-09-949-016-12702
5	23.4	46.8	60592	4	US-09-949-016-13741
6	23.4	46.8	141248	4	US-09-949-016-12241
7	23.4	46.0	512	4	US-09-949-016-12655
8	22.8	45.6	161	2	US-08-613-965-3
9	22.8	45.6	161	2	US-08-918-966-3
10	22.8	45.6	161	3	US-08-921-655-3
11	22.8	45.6	601	4	US-09-949-016-154875
12	22.8	45.6	4084	3	US-08-866-340-1
13	22.8	45.6	4460	3	US-09-103-875-4
14	22.8	45.6	8279	4	US-09-949-016-13478
15	22.8	45.6	62463	4	US-09-949-016-16071
16	22.8	45.6	113100	4	US-09-949-016-12245
17	22.6	45.2	601	4	US-09-949-016-121014
18	22.6	45.2	601	4	US-09-949-016-121960
19	22.6	45.2	601	4	US-09-949-016-121961
20	22.6	45.2	601	4	US-09-949-016-121962
21	22.6	45.2	19389	4	US-09-949-016-15113
22	22.6	45.2	70947	4	US-09-949-016-15165
23	22.4	44.8	52661	4	US-09-949-016-17191
24	22.4	44.8	4403765	3	US-09-103-840A-2
25	22.4	44.8	4411529	3	US-09-103-840A-1
26	22.2	44.4	8879	4	US-09-949-016-11827
27	22.2	44.4	8880	4	US-09-949-016-16003

C 101	21.4	42.8	190078	4	US-09-949-016-17026	Sequence 17026, A	174	20.8	41.6	28958	1	US-08-457-646A-6	Sequence 6, Appli
C 102	21.4	42.8	199945	4	US-09-949-016-15436	Sequence 15436, A	175	20.8	41.6	28958	1	US-08-458-076A-6	Sequence 6, Appli
C 103	21.4	42.8	254405	4	US-09-949-016-14381	Sequence 14381, A	176	20.8	41.6	28958	1	US-08-764-233A-4	Sequence 4, Appli
C 104	21.2	42.4	477	4	US-09-252-991A-11919	Sequence 11919, A	177	20.8	41.6	28958	1	US-08-457-335A-6	Sequence 6, Appli
C 105	21.2	42.4	601	4	US-09-949-016-20846	Sequence 20846, A	178	20.8	41.6	28958	1	US-08-723-214-6	Sequence 6, Appli
C 106	21.2	42.4	601	4	US-09-949-016-110193	Sequence 110193, A	179	20.8	41.6	28958	3	US-09-288-934-6	Sequence 6, Appli
C 107	21.2	42.4	1316	4	US-09-620-312D-767	Sequence 767, App	180	20.8	41.6	30360	4	US-09-949-016-15716	Sequence 15716, A
C 108	21.2	42.4	2034	4	US-09-252-991A-11958	Sequence 11958, A	181	20.8	41.6	41522	4	US-09-949-016-11932	Sequence 11932, A
C 109	21.2	42.4	2082	4	US-09-774-528-367	Sequence 367, App	182	20.8	41.6	41523	4	US-09-949-016-15764	Sequence 15764, A
C 110	21.2	42.4	2493	4	US-09-252-991A-11987	Sequence 11987, A	183	20.8	41.6	41965	4	US-09-949-016-13067	Sequence 13067, A
C 111	21.2	42.4	2991	4	US-09-252-991A-12025	Sequence 12025, A	184	20.8	41.6	49377	1	US-08-764-233A-1	Sequence 1, Appli
C 112	21.2	42.4	3783	3	US-08-506-296B-20	Sequence 20, Appl	185	20.8	41.6	57559	4	US-09-949-016-13077	Sequence 13077, A
C 113	21.2	42.4	5703	3	US-09-280-590A-36	Sequence 36, Appl	186	20.8	41.6	57560	4	US-09-949-016-12536	Sequence 12536, A
C 114	21.2	42.4	5703	4	US-09-892-398-36	Sequence 36, Appl	187	20.8	41.6	61399	4	US-09-949-016-14386	Sequence 14386, A
C 115	21.2	42.4	6751	1	US-07-882-925A-5	Sequence 5, Appli	188	20.8	41.6	87563	3	US-09-453-702B-57	Sequence 57, Appl
C 116	21.2	42.4	6751	1	US-08-184-012C-5	Sequence 5, Appli	189	20.8	41.6	178883	4	US-09-949-016-12733	Sequence 12733, A
C 117	21.2	42.4	8971	4	US-09-949-016-15668	Sequence 15668, A	190	20.8	41.6	178884	4	US-09-949-016-13039	Sequence 13039, A
C 118	21.2	42.4	9038	4	US-09-949-016-13523	Sequence 13523, A	191	20.8	41.6	301828	4	US-09-949-016-13969	Sequence 13969, A
C 119	21.2	42.4	9386	4	US-09-949-016-15739	Sequence 15739, A	192	20.8	41.6	786431	4	US-09-751-389-3	Sequence 3, Appli
C 120	21.2	42.4	12317	4	US-09-949-016-14782	Sequence 14782, A	193	20.6	41.2	563	4	US-09-270-767-1132	Sequence 1132, Ap
C 121	21.2	42.4	20966	4	US-09-984-880-3	Sequence 3, Appli	194	20.6	41.2	563	4	US-09-270-767-16414	Sequence 16414, A
C 122	21.2	42.4	20966	4	US-10-277-032-3	Sequence 3, Appli	195	20.6	41.2	601	4	US-09-949-016-150092	Sequence 150092, A
C 123	21.2	42.4	31623	4	US-09-949-016-15945	Sequence 15945, A	196	20.6	41.2	1856	1	US-08-157-171-3	Sequence 3, Appli
C 124	21.2	42.4	42276	4	US-09-949-016-17218	Sequence 17218, A	197	20.6	41.2	1856	3	US-09-050-159-128	Sequence 128, App
C 125	21.2	42.4	42954	4	US-09-949-016-17123	Sequence 17123, A	198	20.6	41.2	2151	4	US-09-489-039A-3670	Sequence 3670, Ap
C 126	21.2	42.4	42954	4	US-09-949-016-17124	Sequence 17124, A	199	20.6	41.2	2151	1	US-08-146-432-20	Sequence 20, Appl
C 127	21.2	42.4	55674	4	US-09-949-016-12563	Sequence 12563, A	200	20.6	41.2	3113	1	US-08-626-554-2	Sequence 2, Appli
C 128	21.2	42.4	55675	4	US-09-949-016-15706	Sequence 15706, A	201	20.6	41.2	3113	4	US-09-949-016-2417	Sequence 2417, Ap
C 129	21.2	42.4	55827	3	US-09-813-133A-3	Sequence 3, Appli	202	20.6	41.2	5917	3	US-08-692-922-1	Sequence 1, Appli
C 130	21.2	42.4	55827	4	US-10-212-877-3	Sequence 3, Appli	203	20.6	41.2	5917	3	US-08-692-922-1	Sequence 1, Appli
C 131	21.2	42.4	85850	4	US-09-949-016-13424	Sequence 13424, A	204	20.6	41.2	22708	4	US-09-949-016-13216	Sequence 13216, A
C 132	21	42.0	294	2	US-08-675-508-15	Sequence 15, Appl	205	20.6	41.2	22708	4	US-09-949-016-15960	Sequence 15960, A
C 133	21	42.0	601	4	US-09-949-016-66735	Sequence 66735, A	206	20.6	41.2	22708	4	US-09-949-016-15875	Sequence 15875, A
C 134	21	42.0	601	4	US-09-949-016-198638	Sequence 198638, A	207	20.6	41.2	22708	4	US-09-949-016-15875	Sequence 15875, A
C 135	21	42.0	2457	4	US-09-863-859-21	Sequence 21, Appl	208	20.6	41.2	22708	4	US-09-949-016-15875	Sequence 15875, A
C 136	21	42.0	4881	4	US-09-863-859-23	Sequence 23, Appl	209	20.6	41.2	22708	4	US-09-949-016-14159	Sequence 14159, A
C 137	21	42.0	5813	4	US-09-495-714C-1	Sequence 1, Appli	210	20.6	41.2	22708	4	US-09-949-016-14132	Sequence 14132, A
C 138	21	42.0	6112	4	US-09-495-714C-3	Sequence 3, Appli	211	20.4	40.8	33	3	US-08-840-316-62	Sequence 62, Appl
C 139	21	42.0	12298	4	US-09-949-016-13707	Sequence 13707, A	212	20.4	40.8	33	3	US-08-809-523-62	Sequence 62, Appl
C 140	21	42.0	20375	4	US-09-949-016-15927	Sequence 15927, A	213	20.4	40.8	33	3	US-08-471-971-62	Sequence 62, Appl
C 141	21	42.0	28438	4	US-09-820-790B-3	Sequence 3, Appli	214	20.4	40.8	33	3	US-09-402-776-62	Sequence 62, Appl
C 142	21	42.0	41393	4	US-09-949-016-16012	Sequence 16012, A	215	20.4	40.8	33	3	US-08-470-246-62	Sequence 62, Appl
C 143	21	42.0	92304	4	US-09-949-016-15943	Sequence 15943, A	216	20.4	40.8	33	4	US-08-316-765-62	Sequence 62, Appl
C 144	21	42.0	112507	4	US-09-949-016-12420	Sequence 12420, A	217	20.4	40.8	33	4	US-09-724-475-62	Sequence 62, Appl
C 145	21	42.0	112507	4	US-09-949-016-12794	Sequence 12794, A	218	20.4	40.8	33	5	PCT-US93-08849A-62	Sequence 62, Appl
C 146	21	42.0	112508	4	US-09-949-016-15899	Sequence 15899, A	219	20.4	40.8	33	5	PCT-US93-08849A-62	Sequence 62, Appl
C 147	21	42.0	112508	4	US-09-949-016-16590	Sequence 16590, A	220	20.4	40.8	601	4	US-09-949-016-203250	Sequence 203250, A
C 148	21	42.0	160759	4	US-09-949-016-16514	Sequence 16514, A	221	20.4	40.8	601	4	US-09-949-016-203251	Sequence 203251, A
C 149	21	42.0	161652	4	US-09-497-855A-40	Sequence 40, Appl	222	20.4	40.8	601	4	US-09-949-016-203252	Sequence 203252, A
C 150	21	42.0	325034	4	US-09-949-016-14957	Sequence 14957, A	223	20.4	40.8	816	4	US-09-248-796A-565	Sequence 565, App
C 151	21	42.0	389504	4	US-09-949-016-11774	Sequence 11774, A	224	20.4	40.8	838	4	US-09-976-594-56	Sequence 56, Appl
C 152	21	42.0	767877	4	US-09-949-016-12147	Sequence 12147, A	225	20.4	40.8	1024	4	US-09-328-475C-29	Sequence 29, Appl
C 153	21	42.0	767877	4	US-09-949-016-17361	Sequence 17361, A	226	20.4	40.8	1826	4	US-09-793-875-13	Sequence 13, Appl
C 154	20.8	41.6	278	2	US-08-332-766A-42	Sequence 42, Appl	227	20.4	40.8	2243	4	US-09-205-258-144	Sequence 144, App
C 155	20.8	41.6	601	4	US-09-949-016-132241	Sequence 132241, A	228	20.4	40.8	7716	4	US-09-949-016-16614	Sequence 16614, A
C 156	20.8	41.6	601	4	US-09-949-016-122004	Sequence 122004, A	229	20.4	40.8	9308	4	US-09-949-016-13107	Sequence 13107, A
C 157	20.8	41.6	601	4	US-09-949-016-140401	Sequence 140401, A	230	20.4	40.8	9325	4	US-09-902-540-903	Sequence 903, App
C 158	20.8	41.6	601	4	US-09-949-016-142254	Sequence 142254, A	231	20.4	40.8	12151	4	US-09-949-016-16588	Sequence 16588, A
C 159	20.8	41.6	601	4	US-09-949-016-142255	Sequence 142255, A	232	20.4	40.8	12151	2	US-08-734-344-1	Sequence 1, Appli
C 160	20.8	41.6	1146	4	US-08-227-800A-1	Sequence 1, Appli	233	20.4	40.8	13875	4	US-09-949-016-17477	Sequence 17477, A
C 161	20.8	41.6	1146	4	US-08-921-95A-1	Sequence 1, Appli	234	20.4	40.8	27589	4	US-09-949-016-17457	Sequence 17457, A
C 162	20.8	41.6	1427	1	US-09-949-016-3768	Sequence 3768, Ap	235	20.4	40.8	42846	4	US-09-949-016-14117	Sequence 14117, A
C 163	20.8	41.6	1558	2	US-08-467-607-2	Sequence 2, Appli	236	20.4	40.8	46447	4	US-09-949-016-15071	Sequence 15071, A
C 164	20.8	41.6	1558	2	US-08-469-362-2	Sequence 2, Appli	237	20.4	40.8	46447	4	US-09-949-016-15072	Sequence 15072, A
C 165	20.8	41.6	1558	2	US-08-850-392-2	Sequence 2, Appli	238	20.4	40.8	46447	4	US-09-949-016-15073	Sequence 15073, A
C 166	20.8	41.6	2469	4	US-09-949-016-1325	Sequence 1325, Ap	239	20.4	40.8	46447	4	US-09-949-016-15074	Sequence 15074, A
C 167	20.8	41.6	2575	4	US-09-949-016-3434	Sequence 3434, App	240	20.4	40.8	46447	4	US-09-949-016-15075	Sequence 15075, A
C 168	20.8	41.6	2589	4	US-09-949-016-976	Sequence 976, App	241	20.4	40.8	46447	4	US-09-949-016-15076	Sequence 15076, A
C 169	20.8	41.6	16871	4	US-09-949-016-12718	Sequence 12718, A	242	20.4	40.8	46447	4	US-09-949-016-15077	Sequence 15077, A
C 170	20.8	41.6	16874	4	US-09-949-016-15176	Sequence 15176, A	243	20.4	40.8	49313	4	US-09-949-016-15063	Sequence 15063, A
C 171	20.8	41.6	28958	1	US-08-258-261B-6	Sequence 6, Appli	244	20.4	40.8	49313	4	US-09-949-016-15064	Sequence 15064, A
C 172	20.8	41.6	28958	1	US-08-456-837-6	Sequence 6, Appli	245	20.4	40.8	49313	4	US-09-949-016-15065	Sequence 15065, A
C 173	20.8	41.6	28958	1	US-08-457-342-6	Sequence 6, Appli	246	20.4	40.8	49313	4	US-09-949-016-15066	Sequence 15066, A

247	20.4	40.8	49313	4	US-09-949-016-15067	Sequence 15067, A	320	20.2	40.4	35928	4	US-09-668-021-17	Sequence 17, Appl
248	20.4	40.8	49313	4	US-09-949-016-15068	Sequence 15068, A	321	20.2	40.4	43102	4	US-09-949-016-16355	Sequence 16355, A
249	20.4	40.8	49313	4	US-09-949-016-15069	Sequence 15069, A	322	20.2	40.4	90776	4	US-09-949-016-17230	Sequence 17230, A
250	20.4	40.8	49313	4	US-09-949-016-15070	Sequence 15070, A	323	20.2	40.4	97423	4	US-09-949-016-17242	Sequence 12742, A
251	20.4	40.8	49378	4	US-09-949-016-13408	Sequence 13408, A	324	20.2	40.4	97424	4	US-09-949-016-15576	Sequence 15576, A
252	20.4	40.8	56678	4	US-09-949-016-17453	Sequence 17453, A	325	20.2	40.4	106199	4	US-09-949-016-12393	Sequence 12393, A
253	20.4	40.8	59853	4	US-09-949-016-13618	Sequence 13618, A	326	20.2	40.4	107980	4	US-09-949-016-14370	Sequence 14370, A
254	20.4	40.8	59853	4	US-09-949-016-13619	Sequence 13619, A	327	20.2	40.4	112874	4	US-09-949-016-13180	Sequence 13180, A
255	20.4	40.8	59853	4	US-09-949-016-13620	Sequence 13620, A	328	20.2	40.4	119801	4	US-09-949-016-13453	Sequence 13453, A
256	20.4	40.8	59853	4	US-09-949-016-13621	Sequence 13621, A	329	20.2	40.4	222452	4	US-09-949-016-13968	Sequence 12968, A
257	20.4	40.8	59853	4	US-09-949-016-13622	Sequence 13622, A	330	20.2	40.4	253364	4	US-09-949-016-13656	Sequence 12656, A
258	20.4	40.8	59853	4	US-09-949-016-13623	Sequence 13623, A	331	20.2	40.4	253364	4	US-09-949-016-13639	Sequence 13639, A
259	20.4	40.8	59853	4	US-09-949-016-13624	Sequence 13624, A	332	20.2	40.4	268449	4	US-09-949-016-12244	Sequence 17244, A
260	20.4	40.8	59853	4	US-09-949-016-13625	Sequence 13625, A	333	20.2	40.4	411	4	US-09-893-737-149	Sequence 149, Appl
261	20.4	40.8	116966	4	US-09-949-016-17557	Sequence 17557, A	334	20.2	40.0	601	4	US-09-949-016-25546	Sequence 25546, A
262	20.4	40.8	161652	4	US-09-497-855A-40	Sequence 40, Appl	335	20.2	40.0	601	4	US-09-949-016-27869	Sequence 27869, A
263	20.4	40.8	163181	4	US-09-949-016-13730	Sequence 13730, A	336	20.2	40.0	601	4	US-09-949-016-57396	Sequence 57396, A
264	20.4	40.8	219964	4	US-09-949-016-15086	Sequence 15086, A	337	20.2	40.0	601	4	US-09-949-016-77051	Sequence 77051, A
265	20.4	40.8	455726	4	US-09-949-016-11947	Sequence 11947, A	338	20.2	40.0	601	4	US-09-949-016-91975	Sequence 91975, A
266	20.4	40.8	481115	4	US-09-949-016-11950	Sequence 11950, A	339	20.2	40.0	601	4	US-09-949-016-91976	Sequence 91976, A
267	20.4	40.8	786431	4	US-09-751-389-3	Sequence 3, Appl	340	20.2	40.0	601	4	US-09-949-016-91977	Sequence 91977, A
268	20.2	40.4	84	4	US-09-339-922A-10	Sequence 10, Appl	341	20.2	40.0	601	4	US-09-949-016-132175	Sequence 132175, A
269	20.2	40.4	84	4	US-08-791-391A-10	Sequence 10, Appl	342	20.2	40.0	601	4	US-09-949-016-178372	Sequence 178372, A
270	20.2	40.4	84	4	US-09-016-061-10	Sequence 10, Appl	343	20.2	40.0	601	4	US-09-949-016-198551	Sequence 198551, A
271	20.2	40.4	141	4	US-09-513-998C-14029	Sequence 14029, A	344	20.2	40.0	601	4	US-09-949-016-198552	Sequence 198552, A
272	20.2	40.4	303	3	US-08-556-978B-80	Sequence 80, Appl	345	20.2	40.0	756	3	US-08-413-974-3	Sequence 3, Appl
273	20.2	40.4	351	4	US-09-339-922A-1	Sequence 1, Appl	346	20.2	40.0	756	3	US-08-434-418-3	Sequence 3, Appl
274	20.2	40.4	351	4	US-08-791-391A-1	Sequence 1, Appl	347	20.2	40.0	756	3	US-08-433-288-3	Sequence 3, Appl
275	20.2	40.4	351	4	US-09-016-061-1	Sequence 1, Appl	348	20.2	40.0	756	3	US-08-174-739A-3	Sequence 3, Appl
276	20.2	40.4	480	4	US-09-270-767-3030	Sequence 3030, Ap	349	20.2	40.0	756	3	US-08-434-256-3	Sequence 3, Appl
277	20.2	40.4	480	4	US-09-270-767-19312	Sequence 19312, A	350	20.2	40.0	810	3	US-08-413-974-5	Sequence 5, Appl
278	20.2	40.4	601	4	US-09-949-016-44584	Sequence 44584, A	351	20.2	40.0	810	3	US-08-434-418-5	Sequence 5, Appl
279	20.2	40.4	601	4	US-09-949-016-59052	Sequence 59052, A	352	20.2	40.0	810	3	US-08-433-288-5	Sequence 5, Appl
280	20.2	40.4	601	4	US-09-949-016-140999	Sequence 140999, A	353	20.2	40.0	810	3	US-08-174-739A-5	Sequence 5, Appl
281	20.2	40.4	685	1	US-08-451-947-7	Sequence 7, Appl	354	20.2	40.0	810	3	US-08-434-256-5	Sequence 5, Appl
282	20.2	40.4	685	2	US-08-424-826A-7	Sequence 7, Appl	355	20.2	40.0	883	4	US-09-902-540-100	Sequence 100, App
283	20.2	40.4	685	3	US-08-928-694-7	Sequence 7, Appl	356	20.2	40.0	1050	4	US-09-614-221A-349	Sequence 349, App
284	20.2	40.4	685	4	US-08-450-842-7	Sequence 7, Appl	357	20.2	40.0	1123	1	US-07-971-096-3	Sequence 3, Appl
285	20.2	40.4	685	4	US-08-451-390-7	Sequence 7, Appl	358	20.2	40.0	1123	1	US-08-175-096-3	Sequence 3, Appl
286	20.2	40.4	685	5	PCT-US91-06950-7	Sequence 7, Appl	359	20.2	40.0	1332	4	US-09-252-991A-2204	Sequence 2204, Ap
287	20.2	40.4	1190	2	US-08-451-947-9	Sequence 9, Appl	360	20.2	40.0	1455	4	US-09-543-681A-2021	Sequence 2021, Ap
288	20.2	40.4	1190	2	US-08-424-826A-9	Sequence 9, Appl	361	20.2	40.0	2013	4	US-09-252-991A-2365	Sequence 2365, Ap
289	20.2	40.4	1190	3	US-08-928-694-9	Sequence 9, Appl	362	20.2	40.0	2540	4	US-09-023-655-956	Sequence 463, App
290	20.2	40.4	1190	4	US-08-450-842-9	Sequence 9, Appl	363	20.2	40.0	4612	4	US-09-023-655-956	Sequence 956, App
291	20.2	40.4	1190	4	US-08-451-390-9	Sequence 9, Appl	364	20.2	40.0	4992	4	US-09-949-016-14520	Sequence 14520, A
292	20.2	40.4	1190	5	PCT-US91-06950-9	Sequence 9, Appl	365	20.2	40.0	14097	4	US-09-949-016-13661	Sequence 13661, A
293	20.2	40.4	1893	3	US-09-363-970-34	Sequence 34, Appl	366	20.2	40.0	15328	2	US-08-888-497-33	Sequence 33, Appl
294	20.2	40.4	1897	2	US-08-532-547-6	Sequence 6, Appl	367	20.2	40.0	15328	3	US-09-362-230-33	Sequence 33, Appl
295	20.2	40.4	1897	2	US-08-532-547-8	Sequence 8, Appl	368	20.2	40.0	15328	5	PCT-US94-07926-33	Sequence 33, Appl
296	20.2	40.4	1897	3	US-09-019-809-6	Sequence 6, Appl	369	20.2	40.0	16102	4	US-09-949-016-16857	Sequence 16857, A
297	20.2	40.4	1897	3	US-09-019-809-8	Sequence 8, Appl	370	20.2	40.0	16802	4	US-09-949-016-12622	Sequence 12622, A
298	20.2	40.4	1897	3	US-09-471-177-6	Sequence 6, Appl	371	20.2	40.0	16605	4	US-09-949-016-14465	Sequence 14465, A
299	20.2	40.4	1897	3	US-09-471-177-8	Sequence 8, Appl	372	20.2	40.0	20565	4	US-09-949-016-12125	Sequence 12125, A
300	20.2	40.4	2338	1	US-08-425-069-1	Sequence 1, Appl	373	20.2	40.0	20566	4	US-09-949-016-13990	Sequence 13990, A
301	20.2	40.4	2338	2	US-08-317-844B-1	Sequence 17, Appl	374	20.2	40.0	21914	4	US-09-949-016-14627	Sequence 14627, A
302	20.2	40.4	2338	2	US-08-795-868-17	Sequence 17, Appl	375	20.2	40.0	21914	4	US-09-949-016-14732	Sequence 14732, A
303	20.2	40.4	2738	3	US-09-303-069-17	Sequence 17, Appl	376	20.2	40.0	21914	4	US-09-949-016-14733	Sequence 14733, A
304	20.2	40.4	2738	3	US-09-134-250-17	Sequence 17, Appl	377	20.2	40.0	21914	4	US-09-949-016-14733	Sequence 14733, A
305	20.2	40.4	3701	1	US-08-553-279-1	Sequence 1, Appl	378	20.2	40.0	33478	4	US-09-949-016-13615	Sequence 13615, A
306	20.2	40.4	7958	4	US-09-949-016-15531	Sequence 16531, A	379	20.2	40.0	33498	4	US-09-949-016-11982	Sequence 11982, A
307	20.2	40.4	13187	3	US-09-422-936-61	Sequence 61, Appl	380	20.2	40.0	33551	4	US-09-949-016-16666	Sequence 16666, A
308	20.2	40.4	13875	2	US-08-734-344-1	Sequence 1, Appl	381	20.2	40.0	35688	4	US-09-949-016-16873	Sequence 16873, A
309	20.2	40.4	14871	4	US-09-949-016-13013	Sequence 13013, A	382	20.2	40.0	36090	4	US-09-949-016-17424	Sequence 17424, A
310	20.2	40.4	18558	4	US-09-949-016-13688	Sequence 13688, A	383	20.2	40.0	36123	4	US-09-949-016-12402	Sequence 12402, A
311	20.2	40.4	19598	4	US-09-902-540-1143	Sequence 1143, Ap	384	20.2	40.0	53719	4	US-09-918-686-2	Sequence 2, Appl
312	20.2	40.4	23856	4	US-09-949-016-15732	Sequence 15732, A	385	20.2	40.0	60990	4	US-09-949-016-14080	Sequence 14080, A
313	20.2	40.4	24817	4	US-09-949-016-16177	Sequence 16177, A	386	20.2	40.0	65518	4	US-09-949-016-12421	Sequence 12421, A
314	20.2	40.4	26896	4	US-09-949-016-15600	Sequence 15600, A	387	20.2	40.0	65518	4	US-09-949-016-14101	Sequence 14101, A
315	20.2	40.4	30336	4	US-09-949-016-12268	Sequence 12268, A	388	20.2	40.0	92139	4	US-09-918-686-1	Sequence 1, Appl
316	20.2	40.4	30537	4	US-09-949-016-16508	Sequence 16508, A	389	20.2	40.0	119981	4	US-09-949-016-11844	Sequence 11844, A
317	20.2	40.4	35828	3	US-09-449-218D-17	Sequence 17, Appl	390	20.2	40.0	119982	4	US-09-949-016-13606	Sequence 13606, A
318	20.2	40.4	35828	4	US-09-668-529A-17	Sequence 17, Appl	391	20.2	40.0	140925	4	US-09-949-016-11777	Sequence 11777, A
319	20.2	40.4	35828	4	US-09-668-037A-17	Sequence 17, Appl	392	20.2	40.0	140982	4	US-09-949-016-16295	Sequence 16295, A


```
US-09-949-016-12241
; Sequence 12241, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12241
; LENGTH: 141248
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(141248)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12241

Query Match 46.8%; Score 23.4; DB 4; Length 141248;
Best Local Similarity 67.3%; Pred. No. 81;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
DB 121920 CTCTCAACCCAGACCTTTCTCCCTTGGCCCTTGACATCCCAACCTC 121968

RESULT 7
US-09-621-976-2655/c
; Sequence 2655, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2655
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 298...465
US-09-621-976-2655

Query Match 46.0%; Score 23; DB 4; Length 512;
Best Local Similarity 68.1%; Pred. No. 44;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 TACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCA 49
DB 237 TACCCATCAGATGCTTCTATTCCTCTCTTGAATTCATCTTTCA 191

RESULT 8
US-08-613-965-3
; Sequence 3, Application US/08613965
; Patent No. 5916745
```

```
; GENERAL INFORMATION:
; APPLICANT: Robert M. Cook and Ahmed Raafat
; TITLE OF INVENTION: Method For Determination
; OF INVENTION: of Bovine Milk Production Potential
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,965
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-290
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5916745e
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161
; TYPE: nucleotides
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: NO
; HYPOTHETICAL: NO
; ANTI-SENSE: No
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; STRAIN:
; INDIVIDUAL ISOLATE: AR8
; CELL TYPE: mammary gland
; FEATURE:
; NAME/KEY: DNA of bovine mammary gland
; LOCATION: bovine mammary gland
; IDENTIFICATION METHOD: Sequencing
; OTHER INFORMATION: Encodes a portion of
; OTHER INFORMATION: acetyl coenzyme A synthetase
US-08-613-965-3

Query Match 45.6%; Score 22.8; DB 2; Length 161;
Best Local Similarity 66.0%; Pred. No. 42;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
DB 54 CTTACCCACCTGGCAACCAACCCCGTGCAGCTGCCTTACACACCGCCC 103

RESULT 9
US-08-918-966-3
; Sequence 3, Application US/08918966
; Patent No. 5981187
; GENERAL INFORMATION:
; APPLICANT: Robert M. Cook and Ahmed Raafat
; TITLE OF INVENTION: Method For Determination
; OF INVENTION: of Bovine Milk Production Potential
```

;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ian C. McLeod
;; STREET: 2190 Commons Parkway
;; CITY: Okemos
;; STATE: Michigan
;; COUNTRY: USA
;; ZIP: 48864
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
;; MEDIUM TYPE: storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/918,966
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ian C. McLeod
;; REGISTRATION NUMBER: 20,931
;; REFERENCE/DOCKET NUMBER: MSU 4.1-290
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (517) 347-4100
;; TELEFAX: (517) 347-4103
;; TELEX: No. 5981187e
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 161
;; TYPE: nucleotides
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;; MOLECULE TYPE:
;; DESCRIPTION: cDNA
;; HYPOTHETICAL: No
;; ANTI-SENSE: No
;; ORIGINAL SOURCE:
;; ORGANISM: Bovine
;; STRAIN:
;; INDIVIDUAL ISOLATE: AR8
;; CELL TYPE: mammary gland
;; FEATURE:
;; NAME/KEY: DNA of bovine mammary gland
;; LOCATION: bovine mammary gland
;; IDENTIFICATION METHOD: Sequencing
;; OTHER INFORMATION: Encodes a portion of
;; OTHER INFORMATION: acetyl coenzyme A synthetase
;;
US-08-918-966-3

Query Match 45.6%; Score 22.8; DB 2; Length 161;
Best Local Similarity 66.0%; Pred. No. 42;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCCAC 50
DB 54 CTTACCCACCTGGCAACCAACCCCGTGCAGTGCCTTACACACCCGCC 103

RESULT 10
US-08-921-655-3
; Sequence 3, Application US/08921655
; Patent No. 6013496
; GENERAL INFORMATION:
; APPLICANT: Robert M. Cook and Ahmed Raafat
; TITLE OF INVENTION: Method For Determination
; TITLE OF INVENTION: of Bovine Milk Production Potential
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway

;; CITY: Okemos
;; STATE: Michigan
;; COUNTRY: USA
;; ZIP: 48864
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
;; MEDIUM TYPE: storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/921,655
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ian C. McLeod
;; REGISTRATION NUMBER: 20,931
;; REFERENCE/DOCKET NUMBER: MSU 4.1-290
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (517) 347-4100
;; TELEFAX: (517) 347-4103
;; TELEX: No. 6013496e
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 161
;; TYPE: nucleotides
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;; MOLECULE TYPE:
;; DESCRIPTION: cDNA
;; HYPOTHETICAL: No
;; ANTI-SENSE: No
;; ORIGINAL SOURCE:
;; ORGANISM: Bovine
;; STRAIN:
;; INDIVIDUAL ISOLATE: AR8
;; CELL TYPE: mammary gland
;; FEATURE:
;; NAME/KEY: DNA of bovine mammary gland
;; LOCATION: bovine mammary gland
;; IDENTIFICATION METHOD: Sequencing
;; OTHER INFORMATION: Encodes a portion of
;; OTHER INFORMATION: acetyl coenzyme A synthetase
;;
US-08-921-655-3

Query Match 45.6%; Score 22.8; DB 3; Length 161;
Best Local Similarity 66.0%; Pred. No. 42;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCCAC 50
DB 54 CTTACCCACCTGGCAACCAACCCCGTGCAGTGCCTTACACACCCGCC 103

RESULT 11
US-09-949-016-154875/c
; Sequence 154875, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154875
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-154875

Query Match 45.6%; Score 22.8; DB 4; Length 601;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAAATTCCTCCAC 43
DB 289 CTTCGGCCAGATCTAGCCCTCAGGCTGACTTCCATC 248

RESULT 12
US-08-866-340-1
; Sequence 1, Application US/08866340
; Patent No. 6020318
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC
; TITLE OF INVENTION: SEQUENCES AND ANTISENSE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESS: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,340
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Keown, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 106.101.187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-866-340-1

Query Match 45.6%; Score 22.8; DB 3; Length 4084;
Best Local Similarity 66.0%; Pred. No. 73;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAAATTCCTCCAC 50
DB 380 CCTTGGTTCCTGGATGCTTGACTCCCAATTCATCTCTCAAAATGCCAC 429

RESULT 13

US-09-103-875-4
; Sequence 4, Application US/09103875A
; Patent No. 6221849
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC SEQUENCES AND ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; FILE REFERENCE: 106101.194
; CURRENT APPLICATION NUMBER: US/09/103,875A
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: 60/069,865
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 08/866,340
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-103-875-4

Query Match 45.6%; Score 22.8; DB 3; Length 4460;
Best Local Similarity 66.0%; Pred. No. 74;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAAATTCCTCCAC 50
DB 380 CCTTGGTTCCTGGATGCTTGACTCCCAATTCATCTCTCAAAATGCCAC 429

RESULT 14

US-09-949-016-13478/c
; Sequence 13478, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13478
; LENGTH: 8279
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13478

Query Match 45.6%; Score 22.8; DB 4; Length 8279;
Best Local Similarity 66.0%; Pred. No. 82;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAAATTCCTCCAC 50
DB 1669 CTGCGCCCCCAGCTCTCCCTTCCCTATTCATTCCTCCATCCCATCCCC 1620

RESULT 15

US-09-949-016-16071/c
; Sequence 16071, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16071
; LENGTH: 62463
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16071

Query Match          45.6%; Score 22.8; DB 4; Length 62463;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      2 CTACCCACCTGGATCCTTGCCCGCCCTTGAATTCCTCCAA 43
Db      55282 CTTCGCCGCGAGGATCCTAGCCTCAGGCGCTGACTTCCCATC 55241

RESULT 16
US-09-949-016-12245/C
; Sequence 12245, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12245
; LENGTH: 113100
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(113100)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12245

Query Match          45.6%; Score 22.8; DB 4; Length 113100;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      1 CCTACCCACCTGGATCCTTGCCCGCCCTTGAATTCCTCCAA 42
Db      41369 CCTCTCCAGCTGGATCCTTCCCTTACACTTATTTACCAA 41328

RESULT 17
US-09-949-016-121014
; Sequence 121014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121014
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-121014

Query Match          45.2%; Score 22.6; DB 4; Length 601;
Best Local Similarity 68.9%; Pred. No. 62;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      4 ACCACCCCTGGATCCTTGCCCGCCCTTGAATTCCTCCAACTCC 48
Db      321 ACACACCTGGACCTTGGCCCACTCTGCGCGCCCGCCCTCC 365

RESULT 18
US-09-949-016-121960
; Sequence 121960, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121960
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-121960

Query Match          45.2%; Score 22.6; DB 4; Length 601;
Best Local Similarity 68.9%; Pred. No. 62;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1 CCTACCCACCTGGATCCTTGCCCGCCCTTGAATTCCTCCAACTCC 45
Db      241 CCTACACAGCTGGTGTCTACTCCCACTTCATTCCTCCCTGCC 285

RESULT 19
US-09-949-016-121961
; Sequence 121961, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```

/ GENOTYPE: UNKNT.
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 17191
/ LENGTH: 52661
/ TYPE: DNA

```



```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(52661)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17191

Query Match      44.8%; Score 22.4; DB 4; Length 52661;
Best Local Similarity 81.2%; Pred. No. 1.6e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTT 32
Db 40904 CCGTCCCGCCCTGGAGCCAGCCCGCCCTT 40873

RESULT 24
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      44.8%; Score 22.4; DB 3; Length 4403765;
Best Local Similarity 72.5%; Pred. No. 2.2e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATCCCAACCTCCA 49
Db 961745 CCGTTCCTTGGCAGCGCTTGAATTCACGAACCCCA 961784

RESULT 25
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      44.8%; Score 22.4; DB 3; Length 4411529;
Best Local Similarity 72.5%; Pred. No. 2.2e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATCCCAACCTCCA 49
Db 961751 CCGTTCCTTGGCAGCGCTTGAATTCACGAACCCCA 961790

RESULT 26
US-09-949-016-11827/c
; Sequence 11827, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11827
; LENGTH: 8879
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11827

Query Match      44.4%; Score 22.2; DB 4; Length 8879;
Best Local Similarity 69.8%; Pred. No. 1.4e+02;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 CTACCACCTGGATCCTTGGCCCGCCCTTGAATTCCAACC 44
Db 8232 CTGCTCTCCTGGATCCCGCAGCCCACTCTTGGCTTCCCGC 8190

RESULT 27
US-09-949-016-16003/c
; Sequence 16003, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16003
; LENGTH: 8880
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16003

Query Match      44.4%; Score 22.2; DB 4; Length 8880;
Best Local Similarity 69.8%; Pred. No. 1.4e+02;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

5 CCCACCCCTGGATCCTTGGCCCCGCCCTTGAATTCCCAACCCCTCCAC 50

Db 3590 CGCCCCCTGAGTCCAAGCCCCCGCCCCCGTCAGGACCCGCCCCCGCCCC 3545

RESULT 31

```

US-09-949-016-13640
; Sequence 13640, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13640
; LENGTH: 26136
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13640

```

Query Match 44.0%; Score 22; DB 4; Length 26136;
Best Local Similarity 67.4%; Pred. No. 1.9e+02;
Matches 31; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy	1	CCTACCCACCGCTGGATCCTTGTCGCCGCGCCCTTGAATTCCCAACCCT	46
Dβ	21131	CCAACTCCCTTGATTCCTCCACCCCCTTTGAACCTCTCTCCCCCT	21176

RESULT 32

```

US-09-949-016-12770
; Sequence 12770, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12770
; LENGTH: 77661
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(77661)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12770

```

Query Match 44.0%; Score 22; DB 4; Length 77661;
Best Local Similarity 67.4%; Pred. No. 2.3e+02;
Matches 31; Conservative 0; Mismatches 15; Indels 0

Qy 5 CCCACCCCTGGATCCTTGCCCGCCCTTGAAATCCCAACCTCCAC 50

Dp 56153 CCTCCGCCCTTCTCTCCCTCCGCCCTTCTTCCCTCCCTCCCTCCC 56

RESULT 33

```

US-09-949-016-13751
; Sequence 13751, Application US/09949016
; Patent NO. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CUC001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13751
; LENGTH: 77663
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(77663)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13751

```

Query Match 44.0%; Score 22; DB 4; Length 77663;
Best Local Similarity 67.4%; Pred. No. 2.3e+02;
Matches 31; Conservative 0; Mismatches 15; Indels 0

Qy 5 CCCACCTGGATCCTTGGCCCGCCCTTGAAATTCCTCAACCTCTCCAC 50

Db 56153 CCTCCGCCCTTTCCTTCCCTCCCGCCCTTCTTCCCTCCCTCCCTCC 56198

RESULT 34

```

US-09-949-016-17235/c
; Sequence 17235, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17235
; LENGTH: 95122
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17235

```

Query Match 44.0%; Score 22; DB 4; Length 95122;
Best Local Similarity 67.4%; Pred. No. 2.4e+02;
Matches 31; Conservative 0; Mismatches 15; Indels 0

Qy 5 CCGACCTGGATCCTTGTCCTCCCGCCCTTGAAATCCCAACCTCCAC 50
|||||
D6 35275 CCGAGCTGAAGCTATAAACCGAGCTGGCACTCCCAACCTTACC 35230
|||||

```

STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 44.0% Score 22; DB 4; Length 1830121;
Best Local Similarity 73.7%; Pred No. 3.2e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 3 TACCACCTGGATCCTTGGCCCGCCCTTGAATTCCC 40
Db 113719 TACCACACTGGTGTTTTACCCGCCCCCTGTAAACCC 113756

RESULT 37
US-09-949-016-28256/c
Sequence 28256, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ.ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28256
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-28256

Query Match 43.6% Score 21.8; DB 4; Length 601;
Best Local Similarity 65.3%; Pred No. 1.2e+02;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCACACCTCCAC 50

```

```
Db      197  CTCCACACGAGGCTCTGCTTCGACTTTTGCTTCCCAAGCCTCCCC 149

RESULT 38
US-09-949-016-61291/c
; Sequence 61291, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61291
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-61291

Query Match      43.6%; Score 21.8; DB 4; Length 601;
Best Local Similarity 65.3%; Pred. No. 1.2e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      2  CTACCCACCTGGATCCTTGCCCGCCCTTGGAATTCCTCCAACTCCAC 50
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      197  CTCCACACGAGGCTCTGCTTCGACTTTTGCTTCCCAAGCCTCCCC 149

RESULT 39
US-09-949-016-110818/c
; Sequence 110818, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110818
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-110818

Query Match      43.6%; Score 21.8; DB 4; Length 601;
Best Local Similarity 65.3%; Pred. No. 1.2e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      2  CTACCCACCTGGATCCTTGCCCGCCCTTGGAATTCCTCCAACTCCAC 50
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      197  CTCCACACGAGGCTCTGCTTCGACTTTTGCTTCCCAAGCCTCCCC 149

RESULT 40
US-09-949-016-110979/c
; Sequence 110979, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110979
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-110979

Query Match      43.6%; Score 21.8; DB 4; Length 601;
Best Local Similarity 65.3%; Pred. No. 1.2e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      2  CTACCCACCTGGATCCTTGCCCGCCCTTGGAATTCCTCCAACTCCAC 50
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      50  CTAACCAACCCCACTTCCATCCAGCCCTTGCCCAATCTATTCTCCAC 2

RESULT 41
US-09-949-016-111140/c
; Sequence 111140, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111140
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-111140

Query Match      43.6%; Score 21.8; DB 4; Length 601;
Best Local Similarity 65.3%; Pred. No. 1.2e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      2  CTACCCACCTGGATCCTTGCCCGCCCTTGGAATTCCTCCAACTCCAC 50
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      50  CTAACCAACCCCACTTCCATCCAGCCCTTGCCCAATCTATTCTCCAC 2

RESULT 42
US-09-949-016-111301/c
; Sequence 111301, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 111301

; LENGTH: 601

; TYPE: DNA

; ORGANISM: Human

; US-09-949-016-111301

Query Match 43.6%; Score 21.8; DB 4; Length 601;

Best Local Similarity 65.3%; Pred. No. 1.2e+02;

Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCCAC 50

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 50 CTACCCACCCCACTTCCATTCCTCCAGCCCTTGCCTCCCAATCTATCTCCAC 2

RESULT 43

US-09-461-325-88/c

; Sequence 88, Application US/09461325A

; Patent No. 6475753

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 94 Human Secreted Proteins

; FILE REFERENCE: PZ029P1

; CURRENT APPLICATION NUMBER: US/09/461,325A

; CURRENT FILING DATE: 1999-12-14

; EARLIER APPLICATION NUMBER: PCT/US99/13418

; EARLIER FILING DATE: 1999-06-15

; EARLIER APPLICATION NUMBER: 60/089,507

; EARLIER FILING DATE: 1998-06-16

; EARLIER APPLICATION NUMBER: 60/089,508

; EARLIER FILING DATE: 1998-06-16

; EARLIER APPLICATION NUMBER: 60/089,509

; EARLIER FILING DATE: 1998-06-16

; EARLIER APPLICATION NUMBER: 60/089,510

; EARLIER FILING DATE: 1998-06-16

; EARLIER APPLICATION NUMBER: 60/090,112

; EARLIER FILING DATE: 1998-06-22

; EARLIER APPLICATION NUMBER: 60/090,113

; EARLIER FILING DATE: 1998-06-22

; NUMBER OF SEQ ID NOS: 532

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 88

; LENGTH: 699

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (661)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (694)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (696)

; OTHER INFORMATION: n equals a,t,g, or c

; US-09-461-325-88

Query Match

Best Local Similarity 43.6%; Score 21.8; DB 4; Length 699;

Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9 CCCTGGATCTTGGCCCGCCCTTGAATTCCTCCACCCCTCCA 49

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 292 CTCGGATTCCATCCCAACCCCTTCCATACACACCCCTCCA 252

RESULT 44

US-10-012-542-88/c

; Sequence 88, Application US/10012542

; Patent No. 6627741

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 94 Human Secreted Proteins

; FILE REFERENCE: PZ029P1

; CURRENT APPLICATION NUMBER: US/10/012,542

; CURRENT FILING DATE: 2001-12-12

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22

; NUMBER OF SEQ ID NOS: 532

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 88

; LENGTH: 699

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (661)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (694)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (696)

; OTHER INFORMATION: n equals a,t,g, or c

; US-10-012-542-88

Query Match

Best Local Similarity 43.6%; Score 21.8; DB 4; Length 699;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9 CCCTGGATCTTGGCCCGCCCTTGAATTCCTCCACCCCTCCA 49

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 292 CTCGGATTCCATCCCAACCCCTTCCATACACACCCCTCCA 252

RESULT 45

US-10-115-123-88/c

; Sequence 88, Application US/10115123

; Patent No. 6774216

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 94 Human Secreted Proteins

; FILE REFERENCE: PZ029G30AP1D2

; CURRENT APPLICATION NUMBER: US/10/115,123

; CURRENT FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: PCT/US99/13418

; PRIOR FILING DATE: 1999-06-15

; PRIOR APPLICATION NUMBER: 60/089,507

; PRIOR FILING DATE: 1998-06-16

; PRIOR APPLICATION NUMBER: 60/089,508

```
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 88
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (661)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (694)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (696)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-115-123-88

Query Match          43.6%; Score 21.8; DB 4; Length 699;
Best Local Similarity 70.7%; Pred. No. 1.2e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9 CCTCGATCTTGGCCCGCCCTTGGAATTCACACCTCCA 49
Db 292 CTCGGATTCATCCCACTTTCATACACACCTCCA 252

RESULT 46
US-09-949-016-17291/c
; Sequence 17291, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17291
; LENGTH: 16782
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-17291

Query Match          43.6%; Score 21.8; DB 4; Length 16782;
Best Local Similarity 65.3%; Pred. No. 2.1e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 CTACCCACCTCGATCTTGGCCCGCCCTTGGAATTCACACCTCCA 50
Db 1399 CTCGCCACCCCTACTCTCCCAACCCCTACTCTCCCAACCCCTCA 1351

RESULT 47
US-09-949-016-14575
; Sequence 14575, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14575
; LENGTH: 32278
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14575

Query Match          43.6%; Score 21.8; DB 4; Length 32278;
Best Local Similarity 65.3%; Pred. No. 2.4e+02;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACCCACCTCGATCTTGGCCCGCCCTTGGAATTCACACCTCCA 49
Db 452 CTCGCCCAACCCACATCTTACCCATCCCTAGTATCTCTCTTCCA 500

RESULT 48
US-09-952-060-31/c
; Sequence 31, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 36620
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA encoding pMRKADSHIV-1 nef,
; OTHER INFORMATION: noncoding
; US-09-952-060-31

Query Match          43.6%; Score 21.8; DB 4; Length 36620;
Best Local Similarity 70.7%; Pred. No. 2.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

QY 5 CCCACCCCTGGATCCTTGCCCGCCCTTGGAATTCACCAACC 45
Db 34521 CGCGCCCGCACCTTGCCCGCCCACTGCATCCAAATCC 34481

RESULT 49

US-09-952-060-26/c
; Sequence 26, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 37474
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA encoding pMRKAD5 HIV-1 gag, noncoding
US-09-952-060-26

Query Match 43.6%; Score 21.8; DB 4; Length 37474;
Best Local Similarity 70.7%; Pred. No. 2.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 CCCACCCCTGGATCCTTGCCCGCCCTTGGAATTCACCAACC 45
Db 35386 CGCGCCCGCACCTTGCCCGCCCACTGCATCCAAATCC 35346

RESULT 50

US-09-952-060-29/c
; Sequence 29, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814

; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 38519
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA encoding pMRKADSHIV-1 pol,
; OTHER INFORMATION: noncoding
US-09-952-060-29

Query Match 43.6%; Score 21.8; DB 4; Length 38519;
Best Local Similarity 70.7%; Pred. No. 2.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 CCCACCCCTGGATCCTTGCCCGCCCTTGGAATTCACCAACC 45
Db 36420 CGCGCCCGCACCTTGCCCGCCCACTGCATCCAAATCC 36380

Search completed: April 25, 2005, 15:10:14
Job time : 180 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:16:14 ; Search time 624.138 Seconds
(without alignments)
486.355 Million cell updates/sec

Title: US-10-010-476-13

Perfect score: 50

Sequence: 1 CCTACCACCTGATCCTT.....TTGAATTCACACCTCCAC 50

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	50	15	US-10-010-476-13
2	30	60.0	30	15	US-10-010-476-9
3	30	60.0	30	15	US-10-010-476-10
4	26.6	53.2	832	17	US-10-424-599-126243
5	26.6	53.2	1051	11	US-09-876-143-1396
6	26.4	52.8	22444	17	US-10-052-482-97
7	25.2	50.4	422	9	US-09-864-761-16101
8	25.2	50.4	422	9	US-09-864-761-32594
9	25.2	50.4	606	13	US-10-027-632-182868
10	25.2	50.4	606	13	US-10-027-632-182869
11	25.2	50.4	606	13	US-10-027-632-304434

12	25.2	50.4	606	17	US-10-027-632-182868	Sequence 182868,
13	25.2	50.4	606	17	US-10-027-632-182869	Sequence 182869,
14	25.2	50.4	606	17	US-10-027-632-304434	Sequence 304434,
C 15	25.2	50.4	608	13	US-10-027-632-93467	Sequence 93467, A
C 16	25.2	50.4	608	13	US-10-027-632-93467	Sequence 93467, A
17	25.2	50.4	4562	18	US-10-664-705-74	Sequence 74, Appl
C 18	25.0	50.0	814	13	US-10-027-632-132091	Sequence 132091,
C 19	25.0	50.0	814	17	US-10-027-632-132091	Sequence 132091,
C 20	25.0	223556	13	US-10-087-192-394	Sequence 394, App	
C 21	24.4	48.8	6960	15	US-10-311-455-2371	Sequence 2371, App
C 22	24.4	48.8	109977	18	US-10-719-993-6870	Sequence 6870, App
C 23	24.2	48.4	650	13	US-10-027-632-223927	Sequence 223927,
C 24	24.2	48.4	650	17	US-10-027-632-223927	Sequence 223927,
C 25	24.4	48.0	1099	13	US-10-027-632-101090	Sequence 101090,
C 26	24.4	48.0	1099	13	US-10-027-632-101091	Sequence 101091,
C 27	24.4	48.0	1099	13	US-10-027-632-101092	Sequence 101092,
C 28	24.4	48.0	1099	17	US-10-027-632-101090	Sequence 101090,
C 29	24.4	48.0	1099	17	US-10-027-632-101091	Sequence 101091,
C 30	24.4	48.0	1099	17	US-10-027-632-101092	Sequence 101092,
C 31	24.4	48.0	1389	17	US-10-282-122A-33497	Sequence 33497, A
C 32	24.4	48.0	5654	15	US-10-311-455-1848	Sequence 1848, App
C 33	24.4	48.0	5654	17	US-10-221-714A-346	Sequence 346, App
C 34	23.6	47.2	484	18	US-10-425-115-36994	Sequence 36994, A
C 35	23.4	46.8	201	19	US-10-741-600-72568	Sequence 72568, A
C 36	23.4	46.8	291	18	US-10-425-115-179997	Sequence 179997,
C 37	23.4	46.8	349	18	US-10-437-963-1939	Sequence 1939, App
C 38	23.4	46.8	682	17	US-10-424-599-79977	Sequence 79977, A
C 39	23.4	46.8	800	13	US-10-027-632-155300	Sequence 155300,
C 40	23.4	46.8	800	17	US-10-027-632-155300	Sequence 155300,
41	23.4	46.8	826	13	US-10-027-632-155023	Sequence 155023,
42	23.4	46.8	826	13	US-10-027-632-155024	Sequence 155024,
43	23.4	46.8	826	13	US-10-027-632-155023	Sequence 155023,
44	23.4	46.8	826	13	US-10-027-632-155024	Sequence 155024,
C 45	23.4	46.8	38146	18	US-10-719-993-6860	Sequence 6860, App
C 46	23.4	46.8	52128	18	US-10-719-993-7051	Sequence 7051, App
C 47	23.4	46.8	125560	19	US-10-741-600-18002	Sequence 18002, A
C 48	23.2	46.4	593	18	US-10-425-115-53962	Sequence 53962, A
C 49	23.2	46.0	229	18	US-10-437-963-10623	Sequence 10623, A
C 50	23.2	46.0	733	18	US-10-767-701-8694	Sequence 8694, App
C 51	22.8	45.6	211	9	US-09-880-107-1199	Sequence 1199, App
C 52	22.8	45.6	417	18	US-10-437-963-25008	Sequence 25008, A
C 53	22.8	45.6	478	17	US-10-424-599-124806	Sequence 124806, A
C 54	22.8	45.6	527	18	US-10-425-115-44151	Sequence 44151, A
C 55	22.8	45.6	624	13	US-10-027-632-203018	Sequence 203018,
C 56	22.8	45.6	624	17	US-10-027-632-203018	Sequence 203018,
C 57	22.8	45.6	632	13	US-10-027-632-17739	Sequence 17739, A
C 58	22.8	45.6	632	17	US-10-027-632-17739	Sequence 17739, A
C 59	22.8	45.6	655	18	US-10-363-345A-8231	Sequence 8231, App
C 60	22.8	45.6	655	18	US-10-363-345A-8232	Sequence 8232, App
C 61	22.8	45.6	655	19	US-10-363-483A-8231	Sequence 8231, App
C 62	22.8	45.6	655	19	US-10-363-483A-8232	Sequence 8232, App
C 63	22.8	45.6	945	10	US-09-764-891-7021	Sequence 7021, App
C 64	22.8	45.6	963	18	US-10-437-963-66726	Sequence 66726, A
C 65	22.8	45.6	11838	10	US-09-764-891-7020	Sequence 7020, App
C 66	22.8	45.6	84410	18	US-10-722-281-747	Sequence 747, App
C 67	22.8	45.6	143239	18	US-10-323-860-546	Sequence 546, App
C 68	22.6	45.2	435	10	US-09-918-995-5102	Sequence 5102, App
C 69	22.6	45.2	676	17	US-10-424-599-87907	Sequence 87907, A
C 70	22.6	45.2	37459	17	US-10-085-117-103	Sequence 103, App
C 71	22.4	44.8	498	18	US-10-425-115-12386	Sequence 12386, A
C 72	22.4	44.8	508	13	US-10-027-632-48595	Sequence 48595, A
C 73	22.4	44.8	508	13	US-10-027-632-295183	Sequence 295183,
C 74	22.4	44.8	508	17	US-10-027-632-48595	Sequence 48595, A
C 75	22.4	44.8	508	17	US-10-027-632-295183	Sequence 295183,
C 76	22.4	44.8	560	13	US-10-027-632-283068	Sequence 283068,
C 77	22.4	44.8	560	17	US-10-027-632-283068	Sequence 283068,
C 78	22.4	44.8	703	18	US-10-363-345A-10463	Sequence 10463, A
C 79	22.4	44.8	703	18	US-10-363-345A-10463	Sequence 10463, A
C 80	22.4	44.8	703	19	US-10-363-483A-10463	Sequence 10463, A
C 81	22.4	44.8	703	19	US-10-363-483A-10463	Sequence 10463, A
82	22.4	44.8	760	13	US-10-027-632-169066	Sequence 169066,
83	22.4	44.8	760	17	US-10-027-632-169066	Sequence 169066,
84	22.4	44.8	1222	17	US-10-425-114-28702	Sequence 28702, A

c 85	22.4	44.8	1240	17	US-10-425-114-865	Sequence 865, App	158	21.8	43.6	665	18	US-10-363-345A-12356	Sequence 12356, A
c 86	22.4	44.8	1319	18	US-10-425-115-99334	Sequence 99334, A	c 159	21.8	43.6	665	19	US-10-363-483A-12355	Sequence 12355, A
c 87	22.4	44.8	1585	18	US-10-739-930-3649	Sequence 3649, App	c 160	21.8	43.6	665	19	US-10-363-483A-12356	Sequence 12356, A
c 88	22.4	44.8	2330	17	US-10-094-749-554	Sequence 554, App	c 161	21.8	43.6	699	14	US-10-012-542-88	Sequence 88, Appl
c 89	22.4	44.8	5299	14	US-10-270-333-73	Sequence 73, Appl	c 162	21.8	43.6	699	14	US-10-115-123-88	Sequence 88, Appl
c 90	22.4	44.8	7148	14	US-10-239-676-8	Sequence 8, Appl	c 163	21.8	43.6	786	18	US-10-437-963-49716	Sequence 49716, A
c 91	22.4	44.8	7148	14	US-10-239-676-8	Sequence 8, Appl	c 164	21.8	43.6	836	17	US-10-424-599-64529	Sequence 64529, A
c 92	22.4	44.8	80959	9	US-09-858-546-3	Sequence 16, Appl	c 165	21.8	43.6	836	17	US-10-027-632-31420	Sequence 31420, A
c 93	22.4	44.8	231914	19	US-10-741-600-17654	Sequence 3, Appl	c 166	21.8	43.6	1177	13	US-10-027-632-31420	Sequence 31420, A
c 94	22.2	44.4	455	18	US-10-425-115-142707	Sequence 17654, A	c 167	21.8	43.6	1177	17	US-10-027-632-31420	Sequence 31420, A
c 95	22.2	44.4	775	18	US-10-425-115-51726	Sequence 142707, A	c 168	21.8	43.6	1803	14	US-10-188-359-202	Sequence 13052, A
c 96	22.2	44.4	1102	17	US-10-398-221-1434	Sequence 1434, App	c 169	21.8	43.6	2088	17	US-10-104-047-660	Sequence 660, App
c 97	22.2	44.4	28965	17	US-10-292-798-1169	Sequence 1169, App	c 170	21.8	43.6	2170	17	US-10-156-995-58	Sequence 58, Appl
c 98	22.2	44.4	191597	18	US-10-719-993-6802	Sequence 6802, App	c 171	21.8	43.6	2170	17	US-10-156-995-63	Sequence 63, Appl
c 99	22.4	44.0	22	15	US-10-010-476-14	Sequence 14, Appl	c 172	21.8	43.6	2170	17	US-10-156-995-64	Sequence 64, Appl
c 100	22.4	44.0	201	18	US-10-741-601-17985	Sequence 17985, A	c 173	21.8	43.6	2240	17	US-10-156-995-59	Sequence 59, Appl
c 101	22.4	44.0	201	19	US-10-741-600-49108	Sequence 49108, A	c 174	21.8	43.6	2240	17	US-10-188-359-4	Sequence 4, Appl
c 102	22.4	44.0	340	18	US-10-425-115-176531	Sequence 176531, A	c 175	21.8	43.6	2240	17	US-10-188-359-202	Sequence 202, App
c 103	22.4	44.0	472	18	US-10-425-115-33965	Sequence 33965, A	c 176	21.8	43.6	2240	17	US-09-823-245A-438	Sequence 438, App
c 104	22.4	44.0	541	13	US-10-027-632-158918	Sequence 83965, A	c 177	21.8	43.6	6306	14	US-10-239-676-223	Sequence 223, App
c 105	22.4	44.0	541	17	US-10-027-632-158918	Sequence 158918, A	c 178	21.8	43.6	6306	15	US-10-240-453-331	Sequence 331, App
c 106	22.4	44.0	584	16	US-10-029-386-2869	Sequence 2869, App	c 179	21.8	43.6	9432	10	US-09-942-310-1	Sequence 1, Appl
c 107	22.4	44.0	675	13	US-10-027-632-140620	Sequence 140620, App	c 180	21.8	43.6	9432	14	US-10-209-737-1	Sequence 1, Appl
c 108	22.4	44.0	675	17	US-10-027-632-140620	Sequence 140620, A	c 181	21.8	43.6	9432	14	US-10-712-363-1	Sequence 1, Appl
c 109	22.4	44.0	1068	18	US-10-767-701-8396	Sequence 8396, App	c 182	21.8	43.6	9609	19	US-10-209-737-2	Sequence 2, Appl
c 110	22.4	44.0	1089	13	US-10-027-632-264761	Sequence 264761, A	c 183	21.8	43.6	9609	14	US-10-635-780-4	Sequence 4, Appl
c 111	22.4	44.0	1089	17	US-10-027-632-264761	Sequence 264761, A	c 184	21.8	43.6	13677	17	US-10-712-363-6	Sequence 6, Appl
c 112	22.4	44.0	1372	15	US-10-017-161-2187	Sequence 2187, App	c 185	21.8	43.6	13677	19	US-10-712-363-1	Sequence 1, Appl
c 113	22.4	44.0	1372	17	US-10-292-798-1833	Sequence 1833, App	c 186	21.8	43.6	13677	19	US-10-712-363-1	Sequence 1, Appl
c 114	22.4	44.0	1377	17	US-10-282-132A-31029	Sequence 31029, A	c 187	21.8	43.6	13677	19	US-10-712-363-1	Sequence 1, Appl
c 115	22.4	44.0	1389	9	US-09-815-245-6968	Sequence 6868, App	c 188	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 116	22.4	44.0	1389	17	US-10-282-132A-21894	Sequence 21894, A	c 189	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 117	22.4	44.0	1859	17	US-10-291-172-534	Sequence 534, App	c 190	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 118	22.4	44.0	1859	17	US-10-274-639-40	Sequence 40, App	c 191	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 119	22.4	44.0	1892	17	US-10-333-574-40	Sequence 40, Appl	c 192	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 120	22.4	44.0	2211	17	US-10-333-574-40	Sequence 40, Appl	c 193	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 121	22.4	44.0	2211	17	US-10-305-720-1150	Sequence 1150, App	c 194	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 122	22.4	44.0	2231	17	US-10-344-890-1	Sequence 1, Appl	c 195	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 123	22.4	44.0	2291	17	US-10-082-674-1660	Sequence 1660, App	c 196	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 124	22.4	44.0	2291	18	US-10-723-860-1602	Sequence 1602, App	c 197	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 125	22.4	44.0	2569	18	US-10-723-860-1602	Sequence 1602, App	c 198	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 126	22.4	44.0	2774	19	US-10-741-600-7613	Sequence 7613, App	c 199	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 127	22.4	44.0	2896	17	US-10-353-690-87	Sequence 87, Appl	c 200	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 128	22.4	44.0	3123	13	US-10-027-632-116061	Sequence 116061, A	c 201	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 129	22.4	44.0	3123	17	US-10-027-632-116061	Sequence 116061, A	c 202	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 130	22.4	44.0	4164	17	US-10-486-678-17	Sequence 17, Appl	c 203	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 131	22.4	44.0	4164	19	US-10-486-678-17	Sequence 17, Appl	c 204	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 132	22.4	44.0	4177	17	US-10-291-172-158	Sequence 158, App	c 205	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 133	22.4	44.0	4177	17	US-10-221-278-158	Sequence 158, App	c 206	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 134	22.4	44.0	4351	16	US-10-251-196-10	Sequence 10, Appl	c 207	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 135	22.4	44.0	4603	13	US-10-002-600-113	Sequence 113, App	c 208	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 136	22.4	44.0	4690	18	US-10-723-860-8359	Sequence 8359, App	c 209	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 137	22.4	44.0	4839	18	US-10-357-930-30288	Sequence 30288, A	c 210	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 138	22.4	44.0	83445	19	US-10-741-600-17786	Sequence 17786, A	c 211	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 139	22.4	44.0	97247	13	US-10-087-192-688	Sequence 688, App	c 212	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 140	22.4	44.0	166910	17	US-10-292-337-11	Sequence 11, Appl	c 213	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 141	22.4	44.0	176930	18	US-10-741-601-5679	Sequence 5679, App	c 214	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 142	22.4	44.0	176930	19	US-10-741-600-17725	Sequence 17725, A	c 215	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 143	22.4	44.0	1830121	17	US-10-329-670-1	Sequence 1, Appl	c 216	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 144	22.4	44.0	1830121	18	US-10-425-115-50495	Sequence 50495, A	c 217	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 145	21.8	43.6	399	18	US-10-425-115-50495	Sequence 50495, A	c 218	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 146	21.8	43.6	501	9	US-09-783-590-10254	Sequence 10254, A	c 219	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 147	21.8	43.6	513	18	US-10-425-115-161443	Sequence 161443, A	c 220	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 148	21.8	43.6	520	17	US-10-424-599-48729	Sequence 48729, A	c 221	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 149	21.8	43.6	531	9	US-09-864-761-14755	Sequence 14755, A	c 222	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 150	21.8	43.6	533	18	US-10-425-115-165939	Sequence 165939, A	c 223	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 151	21.8	43.6	562	13	US-10-027-632-271685	Sequence 271685, A	c 224	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 152	21.8	43.6	562	17	US-10-027-632-271685	Sequence 271685, A	c 225	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 153	21.8	43.6	593	18	US-10-363-345A-14339	Sequence 14339, A	c 226	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 154	21.8	43.6	593	18	US-10-363-345A-14340	Sequence 14340, A	c 227	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 155	21.8	43.6	593	19	US-10-363-345A-14339	Sequence 14339, A	c 228	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 156	21.8	43.6	593	19	US-10-363-345A-14340	Sequence 14340, A	c 229	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl
c 157	21.8	43.6	665	18	US-10-363-345A-12355	Sequence 12355, A	c 230	21.8	43.6	13677	19	US-10-645-883A-8	Sequence 8, Appl

231	21.6	43.2	36620	19	US-10-636-730-31	Sequence 31, Appl	C 304	21.4	42.8	40060	18	US-10-388-838-68	Sequence 68, Appl
232	21.6	43.2	37474	10	US-09-952-060-26	Sequence 26, Appl	C 305	21.4	42.8	493631	13	US-10-087-192-205	Sequence 205, App
233	21.6	43.2	37474	18	US-10-380-641-26	Sequence 26, Appl	C 306	21.4	42.8	2731748	18	US-10-297-465A-1	Sequence 1, Appl
234	21.6	43.2	37474	18	US-10-645-883A-8	Sequence 8, Appl	C 307	21.4	42.8	3673778	16	US-10-312-841-1	Sequence 1, Appl
235	21.6	43.2	37474	18	US-10-645-187-7	Sequence 7, Appl	C 308	21.4	42.8	9025608	15	US-10-156-761-1	Sequence 1, Appl
236	21.6	43.2	37474	19	US-10-636-730-26	Sequence 26, Appl	C 309	21.2	42.4	213	17	US-10-424-599-36724	Sequence 36724, A
237	21.6	43.2	38519	10	US-09-952-060-29	Sequence 29, Appl	C 310	21.2	42.4	311	11	US-09-864-408A-103	Sequence 103, App
238	21.6	43.2	38519	18	US-10-380-641-29	Sequence 29, Appl	C 311	21.2	42.4	324	11	US-09-864-408A-8817	Sequence 8817, App
239	21.6	43.2	38519	18	US-10-636-730-29	Sequence 29, Appl	C 312	21.2	42.4	326	18	US-10-437-963-23869	Sequence 23869, A
240	21.6	43.2	40046	11	US-09-968-007A-7	Sequence 7, Appl	C 313	21.2	42.4	351	9	US-09-764-869-429	Sequence 429, App
241	21.6	43.2	40046	11	US-09-968-007A-228	Sequence 228, App	C 314	21.2	42.4	351	14	US-10-091-504-429	Sequence 429, App
242	21.6	43.2	40046	11	US-09-968-007A-689	Sequence 689, App	C 315	21.2	42.4	351	17	US-10-227-577-429	Sequence 429, App
243	21.6	43.2	40046	19	US-10-843-641A-6477	Sequence 6477, Ap	C 316	21.2	42.4	418	18	US-10-021-323-7456	Sequence 7456, Ap
244	21.6	43.2	40046	19	US-10-843-641A-6698	Sequence 6698, Ap	C 317	21.2	42.4	420	18	US-10-425-115-96950	Sequence 96950, A
245	21.6	43.2	40046	19	US-10-843-641A-7159	Sequence 7159, Ap	C 318	21.2	42.4	423	18	US-10-425-115-147343	Sequence 147343, A
246	21.6	43.2	66681	13	US-10-322-281-411	Sequence 411, App	C 319	21.2	42.4	436	18	US-10-425-115-162855	Sequence 162855, A
247	21.6	43.2	131576	13	US-10-087-192-1564	Sequence 1564, App	C 320	21.2	42.4	437	18	US-10-425-115-32217	Sequence 32217, A
248	21.6	43.2	203132	18	US-10-322-281-459	Sequence 459, App	C 321	21.2	42.4	443	18	US-10-021-323-10569	Sequence 10569, A
249	21.6	43.2	421609	18	US-10-367-094-122	Sequence 122, App	C 322	21.2	42.4	448	10	US-09-918-995-25392	Sequence 25392, A
250	21.4	42.8	309	17	US-10-424-599-128005	Sequence 128005, A	C 323	21.2	42.4	456	9	US-09-783-590-3630	Sequence 3630, Ap
251	21.4	42.8	406	18	US-10-425-115-144390	Sequence 144390, A	C 324	21.2	42.4	456	10	US-09-918-995-23907	Sequence 23907, A
252	21.4	42.8	463	9	US-09-867-550-391	Sequence 391, App	C 325	21.2	42.4	465	10	US-09-918-995-3081	Sequence 9081, Ap
253	21.4	42.8	482	16	US-10-197-844-19	Sequence 19, Appl	C 326	21.2	42.4	470	10	US-09-918-995-19536	Sequence 19536, A
254	21.4	42.8	487	17	US-10-027-632-230641	Sequence 230641, A	C 327	21.2	42.4	473	10	US-09-918-995-23220	Sequence 23220, A
255	21.4	42.8	487	17	US-10-027-632-230641	Sequence 230641, A	C 328	21.2	42.4	474	17	US-10-424-599-110627	Sequence 110627, A
256	21.4	42.8	496	13	US-10-027-632-289889	Sequence 289889, A	C 329	21.2	42.4	474	18	US-10-767-701-27650	Sequence 27650, A
257	21.4	42.8	496	17	US-10-027-632-289889	Sequence 289889, A	C 330	21.2	42.4	477	18	US-10-425-115-181664	Sequence 181664, A
258	21.4	42.8	539	17	US-10-424-599-80252	Sequence 80252, A	C 331	21.2	42.4	483	17	US-10-424-599-44718	Sequence 44718, A
259	21.4	42.8	613	18	US-10-363-345A-425	Sequence 425, App	C 332	21.2	42.4	503	18	US-10-425-115-63245	Sequence 63245, A
260	21.4	42.8	613	18	US-10-363-345A-426	Sequence 426, App	C 333	21.2	42.4	505	18	US-10-363-345A-34895	Sequence 34895, A
261	21.4	42.8	613	19	US-10-363-483A-425	Sequence 425, App	C 334	21.2	42.4	505	18	US-10-363-345A-34896	Sequence 34896, A
262	21.4	42.8	613	19	US-10-363-483A-426	Sequence 426, App	C 335	21.2	42.4	505	19	US-10-363-483A-34896	Sequence 34896, A
263	21.4	42.8	728	18	US-10-363-345A-2751	Sequence 2751, Ap	C 336	21.2	42.4	506	19	US-10-363-483A-34896	Sequence 34896, A
264	21.4	42.8	728	18	US-10-363-345A-2752	Sequence 2752, Ap	C 337	21.2	42.4	526	18	US-10-363-345A-22645	Sequence 22645, A
265	21.4	42.8	728	19	US-10-363-483A-2751	Sequence 2751, Ap	C 338	21.2	42.4	526	18	US-10-363-345A-22646	Sequence 22646, A
266	21.4	42.8	728	19	US-10-363-483A-2752	Sequence 2752, Ap	C 339	21.2	42.4	526	19	US-10-363-483A-22646	Sequence 22646, A
267	21.4	42.8	765	17	US-10-398-221-1636	Sequence 1636, Ap	C 340	21.2	42.4	526	19	US-10-363-483A-22646	Sequence 22646, A
268	21.4	42.8	773	18	US-10-425-115-53598	Sequence 53598, A	C 341	21.2	42.4	546	18	US-10-363-345A-12775	Sequence 12775, A
269	21.4	42.8	783	9	US-09-778-320-261	Sequence 261, App	C 342	21.2	42.4	546	18	US-10-363-345A-12776	Sequence 12776, A
270	21.4	42.8	783	13	US-09-910-689-261	Sequence 261, App	C 343	21.2	42.4	546	19	US-10-363-483A-12775	Sequence 12775, A
271	21.4	42.8	783	13	US-10-010-742-261	Sequence 261, App	C 344	21.2	42.4	546	18	US-10-363-483A-12776	Sequence 12776, A
272	21.4	42.8	783	18	US-10-714-389-261	Sequence 261, App	C 345	21.2	42.4	565	18	US-10-425-115-47921	Sequence 47921, A
273	21.4	42.8	783	18	US-10-717-296-261	Sequence 261, App	C 346	21.2	42.4	599	14	US-09-777-564-1066	Sequence 1066, Ap
274	21.4	42.8	816	13	US-10-027-632-167839	Sequence 167839, A	C 347	21.2	42.4	599	9	US-10-015-219-1066	Sequence 1066, Ap
275	21.4	42.8	816	13	US-10-027-632-167839	Sequence 167839, A	C 348	21.2	42.4	627	18	US-10-437-963-92015	Sequence 92015, A
276	21.4	42.8	816	15	US-10-156-761-3161	Sequence 3161, Ap	C 349	21.2	42.4	629	18	US-10-021-323-9375	Sequence 9375, Ap
277	21.4	42.8	900	15	US-10-193-002-300	Sequence 300, App	C 350	21.2	42.4	632	18	US-10-437-963-31394	Sequence 31394, A
278	21.4	42.8	900	15	US-10-084-843-305	Sequence 305, App	C 351	21.2	42.4	636	18	US-10-767-701-26654	Sequence 26654, A
279	21.4	42.8	1079	17	US-10-424-599-22511	Sequence 22511, A	C 352	21.2	42.4	650	18	US-10-767-701-20399	Sequence 20399, A
280	21.4	42.8	1410	17	US-10-425-114-6299	Sequence 6299, Ap	C 353	21.2	42.4	653	18	US-10-363-345A-27971	Sequence 27971, A
281	21.4	42.8	1951	9	US-09-736-968A-104	Sequence 104, App	C 354	21.2	42.4	653	18	US-10-363-345A-27972	Sequence 27972, A
282	21.4	42.8	2359	18	US-10-425-115-135791	Sequence 135791, A	C 355	21.2	42.4	653	19	US-10-363-483A-27971	Sequence 27971, A
283	21.4	42.8	3452	9	US-09-735-169A-1	Sequence 1, Appl	C 356	21.2	42.4	682	18	US-10-363-483A-27972	Sequence 27972, A
284	21.4	42.8	3452	9	US-09-735-171A-1	Sequence 1, Appl	C 357	21.2	42.4	682	18	US-10-767-701-23641	Sequence 23641, A
285	21.4	42.8	3452	17	US-10-701-013-1	Sequence 1, Appl	C 358	21.2	42.4	688	13	US-10-027-632-166907	Sequence 166907, A
286	21.4	42.8	3514	17	US-10-172-118-950	Sequence 950, App	C 359	21.2	42.4	688	17	US-10-027-632-166907	Sequence 166907, A
287	21.4	42.8	3514	17	US-10-342-887-950	Sequence 950, App	C 360	21.2	42.4	703	19	US-10-488-056-4	Sequence 4, Appl
288	21.4	42.8	3514	19	US-10-772-636-19	Sequence 19, Appl	C 361	21.2	42.4	707	18	US-10-363-345A-3129	Sequence 3129, Ap
289	21.4	42.8	6172	15	US-10-240-485-125	Sequence 125, App	C 362	21.2	42.4	707	18	US-10-363-345A-3130	Sequence 3130, Ap
290	21.4	42.8	7052	15	US-10-017-161-2411	Sequence 2411, Ap	C 363	21.2	42.4	707	19	US-10-363-483A-3129	Sequence 3129, Ap
291	21.4	42.8	7225	17	US-10-388-934-617	Sequence 617, App	C 364	21.2	42.4	707	19	US-10-363-483A-3130	Sequence 3130, Ap
292	21.4	42.8	7225	18	US-10-283-975A-395	Sequence 395, App	C 365	21.2	42.4	794	18	US-10-425-115-11860	Sequence 11860, A
293	21.4	42.8	10024	9	US-09-880-107-2430	Sequence 2430, Ap	C 366	21.2	42.4	928	18	US-10-363-345A-20865	Sequence 20865, A
294	21.4	42.8	10373	13	US-10-087-192-668	Sequence 668, App	C 367	21.2	42.4	928	18	US-10-363-345A-20866	Sequence 20866, A
295	21.4	42.8	17491	15	US-10-017-161-1995	Sequence 1995, Ap	C 368	21.2	42.4	928	19	US-10-363-483A-20865	Sequence 20865, A
296	21.4	42.8	34722	18	US-10-322-281-700	Sequence 700, App	C 369	21.2	42.4	928	19	US-10-363-483A-20866	Sequence 20866, A
297	21.4	42.8	80105	18	US-10-322-281-223	Sequence 223, App	C 370	21.2	42.4	941	18	US-10-363-345A-37667	Sequence 37667, A
298	21.4	42.8	93483	17	US-10-034-650-49	Sequence 49, Appl	C 371	21.2	42.4	941	18	US-10-363-345A-37668	Sequence 37668, A
299	21.4	42.8	102374	14	US-10-087-192-667	Sequence 667, App	C 372	21.2	42.4	941	19	US-10-363-483A-37667	Sequence 37667, A
300	21.4	42.8	185548	14	US-10-175-523-62	Sequence 62, Appl	C 373	21.2	42.4	941	19	US-10-363-483A-37668	Sequence 37668, A
301	21.4	42.8	301892	17	US-10-428-487-11	Sequence 11, Appl	C 374	21.2	42.4	968	18	US-10-437-963-41539	Sequence 41539, A
302	21.4	42.8	304905	17	US-10-271-416-1	Sequence 1, Appl	C 375	21.2	42.4	987	18	US-10-363-345A-36685	Sequence 36685, A
303	21.4	42.8	310268	18	US-10-367-094-195	Sequence 195, App	C 376	21.2	42.4	987	18	US-10-363-345A-36686	Sequence 36686, A

C 377	21.2	42.4	987	19	US-10-363-483A-36685	Sequence 36685, A	C 450	21.2	42.4	2276	14	US-10-131-813A-9	Sequence 9, Appli
C 378	21.2	42.4	987	17	US-10-363-483A-36686	Sequence 36686, A	C 451	21.2	42.4	2276	14	US-10-131-818A-9	Sequence 9, Appli
C 379	21.2	42.4	1009	17	US-10-424-599-1962	Sequence 1962, Ap	C 452	21.2	42.4	2276	14	US-10-131-823A-9	Sequence 9, Appli
C 380	21.2	42.4	1045	13	US-10-044-090-678	Sequence 678, App	C 453	21.2	42.4	2276	14	US-10-131-824A-9	Sequence 9, Appli
C 381	21.2	42.4	1095	18	US-10-363-345A-22681	Sequence 22681, A	C 454	21.2	42.4	2276	14	US-10-131-830A-9	Sequence 9, Appli
C 382	21.2	42.4	1095	18	US-10-363-345A-22682	Sequence 22682, A	C 455	21.2	42.4	2276	14	US-10-131-837A-9	Sequence 9, Appli
C 383	21.2	42.4	1095	19	US-10-363-483A-22681	Sequence 22681, A	C 456	21.2	42.4	2276	14	US-10-131-872A-9	Sequence 9, Appli
C 384	21.2	42.4	1095	19	US-10-363-483A-22682	Sequence 22682, A	C 457	21.2	42.4	2276	14	US-10-147-500-9	Sequence 9, Appli
C 385	21.2	42.4	1164	18	US-10-363-345A-4463	Sequence 4463, Ap	C 458	21.2	42.4	2276	14	US-10-147-502-9	Sequence 9, Appli
C 386	21.2	42.4	1164	18	US-10-363-345A-4464	Sequence 4464, Ap	C 459	21.2	42.4	2276	14	US-10-147-515-9	Sequence 9, Appli
C 387	21.2	42.4	1164	19	US-10-363-483A-4463	Sequence 4463, Ap	C 460	21.2	42.4	2276	14	US-10-147-517-9	Sequence 9, Appli
C 388	21.2	42.4	1235	9	US-10-363-483A-4464	Sequence 4464, Ap	C 461	21.2	42.4	2276	14	US-10-147-526-9	Sequence 9, Appli
C 389	21.2	42.4	1235	9	US-09-764-865-2083	Sequence 2083, Ap	C 462	21.2	42.4	2276	14	US-10-147-527-9	Sequence 9, Appli
C 390	21.2	42.4	1235	14	US-09-764-865-2084	Sequence 2084, Ap	C 463	21.2	42.4	2276	14	US-10-121-041-9	Sequence 9, Appli
C 391	21.2	42.4	1235	14	US-10-091-504-2083	Sequence 2083, Ap	C 464	21.2	42.4	2276	14	US-10-121-043-9	Sequence 9, Appli
C 392	21.2	42.4	1235	14	US-10-091-504-2084	Sequence 2084, Ap	C 465	21.2	42.4	2276	14	US-10-121-047-9	Sequence 9, Appli
C 393	21.2	42.4	1235	17	US-10-227-577-2083	Sequence 2083, Ap	C 466	21.2	42.4	2276	14	US-10-123-215-9	Sequence 9, Appli
C 394	21.2	42.4	1235	17	US-10-227-577-2084	Sequence 2084, Ap	C 467	21.2	42.4	2276	14	US-10-123-902-9	Sequence 9, Appli
C 395	21.2	42.4	1299	17	US-10-424-599-92672	Sequence 92672, A	C 468	21.2	42.4	2276	14	US-10-123-908-9	Sequence 9, Appli
C 396	21.2	42.4	1316	15	US-10-037-270-767	Sequence 767, App	C 469	21.2	42.4	2276	14	US-10-123-909-9	Sequence 9, Appli
C 397	21.2	42.4	1316	17	US-10-117-722-767	Sequence 767, App	C 470	21.2	42.4	2276	14	US-10-123-910-9	Sequence 9, Appli
C 398	21.2	42.4	1322	19	US-10-874-706-76	Sequence 76, Appli	C 471	21.2	42.4	2276	14	US-10-124-813-9	Sequence 9, Appli
C 399	21.2	42.4	1336	14	US-10-116-255-7	Sequence 7, Appli	C 472	21.2	42.4	2276	14	US-10-124-817-9	Sequence 9, Appli
C 400	21.2	42.4	1361	17	US-10-369-493-44741	Sequence 44741, A	C 473	21.2	42.4	2276	14	US-10-125-922-9	Sequence 9, Appli
C 401	21.2	42.4	1633	17	US-10-104-047-1046	Sequence 1046, Ap	C 474	21.2	42.4	2276	14	US-10-125-924-9	Sequence 9, Appli
C 402	21.2	42.4	1747	18	US-10-363-345A-4855	Sequence 4855, Ap	C 475	21.2	42.4	2276	14	US-10-140-860-9	Sequence 9, Appli
C 403	21.2	42.4	1747	18	US-10-363-345A-4856	Sequence 4856, Ap	C 476	21.2	42.4	2276	14	US-10-142-417-9	Sequence 9, Appli
C 404	21.2	42.4	1747	19	US-10-363-483A-4855	Sequence 4855, Ap	C 477	21.2	42.4	2276	14	US-10-147-519-9	Sequence 9, Appli
C 405	21.2	42.4	1747	19	US-10-363-483A-4856	Sequence 4856, Ap	C 478	21.2	42.4	2276	14	US-10-157-782-9	Sequence 9, Appli
C 406	21.2	42.4	1852	9	US-09-969-853-4	Sequence 4, Appli	C 479	21.2	42.4	2276	14	US-10-152-395-9	Sequence 9, Appli
C 407	21.2	42.4	2037	9	US-09-822-830A-574	Sequence 574, App	C 480	21.2	42.4	2276	14	US-10-125-926A-9	Sequence 9, Appli
C 408	21.2	42.4	2060	17	US-10-108-260A-1072	Sequence 1072, Ap	C 481	21.2	42.4	2276	14	US-10-125-930A-9	Sequence 9, Appli
C 409	21.2	42.4	2082	17	US-10-120-988-367	Sequence 367, App	C 482	21.2	42.4	2276	14	US-10-127-831A-9	Sequence 9, Appli
C 410	21.2	42.4	2089	17	US-10-369-493-26800	Sequence 26800, A	C 483	21.2	42.4	2276	14	US-10-127-837A-9	Sequence 9, Appli
C 411	21.2	42.4	2203	18	US-10-821-273-93	Sequence 93, Appli	C 484	21.2	42.4	2276	14	US-10-127-838B-9	Sequence 9, Appli
C 412	21.2	42.4	2268	17	US-10-470-360-41	Sequence 41, Appli	C 485	21.2	42.4	2276	14	US-10-127-842A-9	Sequence 9, Appli
C 413	21.2	42.4	2276	14	US-10-028-072-9	Sequence 9, Appli	C 486	21.2	42.4	2276	14	US-10-127-843A-9	Sequence 9, Appli
C 414	21.2	42.4	2276	14	US-10-140-808-9	Sequence 9, Appli	C 487	21.2	42.4	2276	14	US-10-127-845A-9	Sequence 9, Appli
C 415	21.2	42.4	2276	14	US-10-121-049-9	Sequence 9, Appli	C 488	21.2	42.4	2276	14	US-10-127-846A-9	Sequence 9, Appli
C 416	21.2	42.4	2276	14	US-10-123-904-9	Sequence 9, Appli	C 489	21.2	42.4	2276	14	US-10-127-848A-9	Sequence 9, Appli
C 417	21.2	42.4	2276	14	US-10-140-470-9	Sequence 9, Appli	C 490	21.2	42.4	2276	14	US-10-127-849A-9	Sequence 9, Appli
C 418	21.2	42.4	2276	14	US-10-175-746-9	Sequence 9, Appli	C 491	21.2	42.4	2276	14	US-10-127-850A-9	Sequence 9, Appli
C 419	21.2	42.4	2276	14	US-10-176-918-9	Sequence 9, Appli	C 492	21.2	42.4	2276	14	US-10-127-851A-9	Sequence 9, Appli
C 420	21.2	42.4	2276	14	US-10-176-921-9	Sequence 9, Appli	C 493	21.2	42.4	2276	14	US-10-128-684A-9	Sequence 9, Appli
C 421	21.2	42.4	2276	14	US-10-137-865-9	Sequence 9, Appli	C 494	21.2	42.4	2276	14	US-10-128-686A-9	Sequence 9, Appli
C 422	21.2	42.4	2276	14	US-10-140-474-9	Sequence 9, Appli	C 495	21.2	42.4	2276	14	US-10-128-690A-9	Sequence 9, Appli
C 423	21.2	42.4	2276	14	US-10-142-431-9	Sequence 9, Appli	C 496	21.2	42.4	2276	14	US-10-128-691A-9	Sequence 9, Appli
C 424	21.2	42.4	2276	14	US-10-143-114-9	Sequence 9, Appli	C 497	21.2	42.4	2276	14	US-10-131-819A-9	Sequence 9, Appli
C 425	21.2	42.4	2276	14	US-10-143-032-9	Sequence 9, Appli	C 498	21.2	42.4	2276	14	US-10-131-829A-9	Sequence 9, Appli
C 426	21.2	42.4	2276	14	US-10-123-262-9	Sequence 9, Appli	C 499	21.2	42.4	2276	14	US-10-131-836A-9	Sequence 9, Appli
C 427	21.2	42.4	2276	14	US-10-142-423-9	Sequence 9, Appli	C 500	21.2	42.4	2276	14	US-10-146-729-9	Sequence 9, Appli
C 428	21.2	42.4	2276	14	US-10-121-050-9	Sequence 9, Appli							
C 429	21.2	42.4	2276	14	US-10-141-755-9	Sequence 9, Appli							
C 430	21.2	42.4	2276	14	US-10-143-032-9	Sequence 9, Appli							
C 431	21.2	42.4	2276	14	US-10-123-108-9	Sequence 9, Appli							
C 432	21.2	42.4	2276	14	US-10-123-236-9	Sequence 9, Appli							
C 433	21.2	42.4	2276	14	US-10-123-261-9	Sequence 9, Appli							
C 434	21.2	42.4	2276	14	US-10-140-921-9	Sequence 9, Appli							
C 435	21.2	42.4	2276	14	US-10-140-928-9	Sequence 9, Appli							
C 436	21.2	42.4	2276	14	US-10-160-498-9	Sequence 9, Appli							
C 437	21.2	42.4	2276	14	US-10-124-824-9	Sequence 9, Appli							
C 438	21.2	42.4	2276	14	US-10-127-825A-9	Sequence 9, Appli							
C 439	21.2	42.4	2276	14	US-10-127-829A-9	Sequence 9, Appli							
C 440	21.2	42.4	2276	14	US-10-127-835A-9	Sequence 9, Appli							
C 441	21.2	42.4	2276	14	US-10-127-839A-9	Sequence 9, Appli							
C 442	21.2	42.4	2276	14	US-10-127-901A-9	Sequence 9, Appli							
C 443	21.2	42.4	2276	14	US-10-128-693A-9	Sequence 9, Appli							
C 444	21.2	42.4	2276	14									
C 445	21.2	42.4	2276	14									
C 446	21.2	42.4	2276	14									
C 447	21.2	42.4	2276	14									
C 448	21.2	42.4	2276	14									
C 449	21.2	42.4	2276	14									

ALIGNMENTS

RESULT 1

US-10-010-476-13

; Sequence 13, Application US/10010476

; Publication No. US20030114402A1

; GENERAL INFORMATION:

; APPLICANT: Reich, No. US20030114402A1bert O.

; Flynn, James

; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5

; NUMBER OF SEQUENCES: 110

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merxant, Gould, Smith, Edell, Welter & Schmidt

; STREET: 11150 Santa Monica Boulevard, Suite 400

; CITY: Los Angeles

; COUNTRY: USA

; ZIP: 90025

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-010-476-13
Query Match 100.0%; Score 50; DB 15; Length 50;
Best Local Similarity 100.0%; Pred. No. 3.2e-08;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCAC 50
Db 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCAC 50
RESULT 2
US-10-010-476-9/c
Sequence 9, Application US/10010476
Publication No. US20030114402A1
GENERAL INFORMATION:
APPLICANT: Reich, No. US20030114402Albert O.
Flynn, James
TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
METHYLTRANSFERASE AND METHODS FOR USE THEREOF
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-010-476-10
Query Match 60.0%; Score 30; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 11 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 40
Db 30 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 1
RESULT 3
US-10-010-476-10
Sequence 10, Application US/10010476
Publication No. US20030114402A1
GENERAL INFORMATION:
APPLICANT: Reich, No. US20030114402Albert O.
Flynn, James
TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
METHYLTRANSFERASE AND METHODS FOR USE THEREOF
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-010-476-10
Query Match 60.0%; Score 30; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 11 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 40
Db 30 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 1

```
Best Local Similarity 100.0%; Score 26.6; DB 17; Length 832;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 40
      |||||
DB 1 CTGGATCCTTGGCCCGCCCTTGAATTCCTCC 30

RESULT 4
US-10-424-599-126243/c
; Sequence 126243, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 126243
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(832)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_85004C.1
US-10-424-599-126243

Query Match 53.2%; Score 26.6; DB 17; Length 832;
Best Local Similarity 71.4%; Pred. No. 4.9;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCACCTCCAC 50
      |||||
DB 507 CTACCCACGCGACGATCCTCTCCCGCCCTGTTTAGTCCTCAACCTCCAC 459

RESULT 5
US-09-876-143-1396/c
; Sequence 1396, Application US/09876143
; Publication No. US20040081958A1
; GENERAL INFORMATION:
; APPLICANT: Infigen Inc.
; APPLICANT: EILERTSEN, KENNETH J.
; APPLICANT: PFISTER-GENSKOW, MARTHA
; APPLICANT: CHILDS, LYNETTE
; APPLICANT: FORSYTHE, TODD
; APPLICANT: BISHOP, MICHAEL D.
; TITLE OF INVENTION: IDENTIFICATION AND USE OF MOLECULAR MARKERS INDICATING
; FILE REFERENCE: 028040-0202
; CURRENT APPLICATION NUMBER: US/09/876,143
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/209,874
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 1744
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1396
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1051)
; OTHER INFORMATION: n is a, c, g, or t
US-09-876-143-1396
```

```
Query Match 53.2%; Score 26.6; DB 11; Length 1051;
Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCACCTCC 48
      |||||
DB 767 CCTCCCNATNAANTTCTTGTCCCNCCCTTNAATTCACCCCTCC 720

RESULT 6
US-10-052-482-97
; Sequence 97, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 22444
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (781)..(1359)
; OTHER INFORMATION: "n" at positions 781 to 1359 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4375)..(5965)
; OTHER INFORMATION: "n" at positions 4375 to 5965 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (18075)..(22444)
; OTHER INFORMATION: "n" at positions 18075 to 22444
US-10-052-482-97

Query Match 52.8%; Score 26.4; DB 17; Length 22444;
Best Local Similarity 75.0%; Pred. No. 4.4;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 7 CACCTGGATCCTTGGCCCGCCCTTGAATTCCTCCACCTCCAC 50
      |||||
DB 9144 CACCTGGTTCCTTGGCCCGCCCTTCCCTCCCTCCCTCCCTCC 9187

RESULT 7
US-09-864-761-16101/c
; Sequence 16101, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```

PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 16101
LENGTH: 422
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049873.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-16101

Query Match 50.4%; Score 25.2; DB 9; Length 422;
Best Local Similarity 78.9%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGATCTTCCCGCCCTTGATTCCTCCCAACCTC 47
DB 159 CCTGTCTCTGACCCCTGTCTGTCATTCCTCCCAACCTC 122

RESULT 8

US-09-864-761-32594/c
Sequence 32594, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32594
LENGTH: 422
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049873.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: NT HIT: g16754525, EVALUE 5.70e-02
OTHER INFORMATION: SWISSPROT HIT: Q9Z7N9, EVALUE 3.70e-01
US-09-864-761-32594

Query Match 50.4%; Score 25.2; DB 9; Length 422;
Best Local Similarity 78.9%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGATCTTCCCGCCCTTGATTCCTCCCAACCTC 47
DB 159 CCTGTCTCTGACCCCTGTCTGTCATTCCTCCCAACCTC 122

RESULT 9

US-10-027-632-182868
Sequence 182868, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182868
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-182868

Query Match
Best Local Similarity 50.4%; Score 25.2; DB 13; Length 606;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATTCCTCCCAACCCCTC 47
Db 28 CCTGTGCTCTGACCCCTGTCTTGCATTCCTCCCAACCCCTC 65

RESULT 10
US-10-027-632-182869
; Sequence 182869, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182869
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-182869

Query Match
Best Local Similarity 50.4%; Score 25.2; DB 13; Length 606;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATTCCTCCCAACCCCTC 47
Db 28 CCTGTGCTCTGACCCCTGTCTTGCATTCCTCCCAACCCCTC 65

RESULT 11
US-10-027-632-304434
; Sequence 304434, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182869
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-182869

Query Match
Best Local Similarity 50.4%; Score 25.2; DB 13; Length 606;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATTCCTCCCAACCCCTC 47
Db 28 CCTGTGCTCTGACCCCTGTCTTGCATTCCTCCCAACCCCTC 65

RESULT 12
US-10-027-632-182868
; Sequence 182868, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182868
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-182868

Query Match
Best Local Similarity 50.4%; Score 25.2; DB 17; Length 606;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATTCCTCCCAACCCCTC 47
Db 28 CCTGTGCTCTGACCCCTGTCTTGCATTCCTCCCAACCCCTC 65

RESULT 13
```



```

US-10-027-632-182869
; Sequence 182869, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182869
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-182869

Query Match          50.4%; Score 25.2; DB 17; Length 606;
Best Local Similarity 78.9%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 CCTGGATCCTTGGCCGCCCTTGAATTCCTCAACCCCTC 47
Db 28 CCTGTGTCCTGACCCCTGCTCTTGCATTCCTCAACCCCTC 65

RESULT 14
US-10-027-632-304434
; Sequence 304434, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 304434
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-304434

```

```
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93467
; LENGTH: 608
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(608)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-93467

Query Match          50.4%; Score 25.2; DB 17; Length 608;
Best Local Similarity 78.9%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 10 CCTGGATCCTTGGCCCGCCCTTGAATTCCTCCCAACCTC 47
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 581 CCTGTGCTCTGACCCCTGTCTTGCATTCCTCAACCTC 544

RESULT 17
US-10-664-705-74
; Sequence 74, Application US/10664705
; Publication No. US20040152107A1
; GENERAL INFORMATION:
; APPLICANT: Altar, Anthony C.
; APPLICANT: Laeng, Pascal
; APPLICANT: Young, Theresa A.
; APPLICANT: Charles, Vinod
; APPLICANT: Bukhman, Yuri
; APPLICANT: Jurata, Linda
; TITLE OF INVENTION: GENE SIGNATURE OF ELECTROSHOCK THERAPY AND METHODS OF USE
; FILE REFERENCE: 03235/100M087-US2
; CURRENT APPLICATION NUMBER: US/10/664,705
; CURRENT FILING DATE: 2003-09-18
; PRIOR APPLICATION NUMBER: US 60/411,718
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/431,882
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 60/479,970
; PRIOR FILING DATE: 2003-06-18
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 4562
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-664-705-74

Query Match          50.4%; Score 25.2; DB 18; Length 4562;
Best Local Similarity 78.9%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCTTGGCCCGCCCTTGAATTC 38
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3178 CCTATCGTCTGGATCTTGGCCCGCCACCTTGAGTTC 3215

RESULT 18
US-10-027-632-132091/c
; Sequence 132091, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

Query Match          50.0%; Score 25; DB 17; Length 814;
Best Local Similarity 75.6%; Pred. No. 18;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132091
; LENGTH: 814
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-132091

Query Match          50.0%; Score 25; DB 13; Length 814;
Best Local Similarity 75.6%; Pred. No. 18;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCTTGGCCCGCCCTTGAATTCCTCA 41
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 CCTCCACCTCTACCTTGGCCCGCCCTTGAATTCCTCA 255

RESULT 19
US-10-027-632-132091/c
; Sequence 132091, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132091
; LENGTH: 814
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-132091

Query Match          50.0%; Score 25; DB 17; Length 814;
Best Local Similarity 75.6%; Pred. No. 18;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```



```

US-10-027-632-223927/c
; Sequence 223927, Application US/10027632
; Publication NO. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223927
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-223927

```

```

Query Match          48.0%; Score 24; DB 13; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0;

Qy      1  CTAACCCACCTGGATCTCTGCCCCGCCCTTCAATTCCCAACCCCTCC 48
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      781  CTTCCCTCCCTGAAGCCCTCCCGGCCATCTGCATCTCCATCCCTTC 734

RESULT 26
US-10-027-632-101091/c
; Sequence 101091, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101091
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101091

```

```

RESULT 25
US-10-027-632-101090/c
; Sequence 101090, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101090
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101090

```

```

Query Match          48.0%; Score 24; DB 13; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0;

Qy  1  CTTACCCACCTGGATCCTTGCCTCCCGCCCTTGAATTCACCAACCTCC 48
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  781  CCTCCCTCCCTGAAGCCCTCCCGGCCACTCTCACTCCCATCCCTTC 734

RESULT 27
US-10-027-632-101092/c
; Sequence 101092, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23

```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101092
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101092

Query Match      48.0%; Score 24; DB 13; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CCTCCCTCCCTGAAGCCCTCCCGGCCCATCTGCATCCCATCCCTTC 734

RESULT 28
US-10-027-632-101090/c
; Sequence 101090, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101090
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101090

Query Match      48.0%; Score 24; DB 17; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CCTCCCTCCCTGAAGCCCTCCCGGCCCATCTGCATCCCATCCCTTC 734

RESULT 29
US-10-027-632-101091/c
; Sequence 101091, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101091
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101091

Query Match      48.0%; Score 24; DB 17; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CCTCCCTCCCTGAAGCCCTCCCGGCCCATCTGCATCCCATCCCTTC 734

RESULT 30
US-10-027-632-101092/c
; Sequence 101092, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101092
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101092

Query Match      48.0%; Score 24; DB 17; Length 1099;
Best Local Similarity 68.8%; Pred. No. 40;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCCCTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CCTCCCTCCCTGAAGCCCTCCCGGCCCATCTGCATCCCATCCCTTC 734
```

```
RESULT 31
US-10-282-122A-33497/c
; Sequence 33497, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33497
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33497

Query Match 48.0%; Score 24; DB 17; Length 1389;
Best Local Similarity 68.8%; Pred. No. 39;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCTGCGCCCTTGAATTCCTCCACCTCC 48
DB 302 CCTTCAACAGCAAACTTGCCTGCGCCCTTGAATTCCTCCACCTCC 255

RESULT 32
US-10-311-455-1848/c
; Sequence 1848, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
```

```
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1848
; LENGTH: 5654
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1848

Query Match 48.0%; Score 24; DB 15; Length 5654;
Best Local Similarity 68.8%; Pred. No. 35;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCTGCGCCCTTGAATTCCTCCACCTCC 48
DB 2367 CCCACCCACCTTACTACTCTCCTCACCACAAATCCCAACCTAC 2320

RESULT 33
US-10-221-714A-346/c
; Sequence 346, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 346
; LENGTH: 5654
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-346

Query Match 48.0%; Score 24; DB 17; Length 5654;
Best Local Similarity 68.8%; Pred. No. 35;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCTGCGCCCTTGAATTCCTCCACCTCC 48
DB 2367 CCCACCCACCTTACTACTCTCCTCACCACAAATCCCAACCTAC 2320

RESULT 34
US-10-425-115-36994/c
; Sequence 36994, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 369326
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_133742C.1
US-10-425-115-36994

Query Match 47.2%; Score 23.6; DB 18; Length 484;
Best Local Similarity 69.6%; Pred. No. 59; Mismatches 14; Indels 0; Gaps 0;
Matches 32; Conservative 0;
QY 5 CCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
DB 481 CCCCCCGGTCTTCGCGCCGGCGTGAGGCCACCCCGCCCC 436

RESULT 35
US-10-741-600-72568/c
; Sequence 72568, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72568
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-72568

Query Match 46.8%; Score 23.4; DB 19; Length 201;
Best Local Similarity 67.3%; Pred. No. 75; Mismatches 16; Indels 0; Gaps 0;
Matches 33; Conservative 0;
QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
DB 57 CTCTCCACCCAGACCTTTTCCTGGCCCTTGACATCCCAACACCTC 9

RESULT 36
US-10-425-115-179997/c
; Sequence 179997, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 179997
; LENGTH: 291
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(291)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_95737C.1
US-10-425-115-179997

Query Match 46.8%; Score 23.4; DB 18; Length 291;
Best Local Similarity 67.3%; Pred. No. 72; Mismatches 16; Indels 0; Gaps 0;
Matches 33; Conservative 0;
QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCCA 49
DB 258 CCCACCCACCTAGTCCAGTCCAAGCCCTGGGGCTACCAACCTCCCA 210

RESULT 37
US-10-437-963-1939
; Sequence 1939, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 1939
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(349)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101759C.1
US-10-437-963-1939

Query Match 46.8%; Score 23.4; DB 18; Length 349;
Best Local Similarity 65.2%; Pred. No. 71; Mismatches 16; Indels 0; Gaps 0;
Matches 30; Conservative 0;
QY 5 CCCACCTGGATCCTTGGCCCGCCCTTGAATTCCTCAACCTCCAC 50
DB 215 CCNCCCGGCGCCGTCCTTGGNATTCCTCAACCCGGGAC 260

RESULT 38
US-10-424-599-79977/c
; Sequence 79977, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 79977

```
; LENGTH: 682
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43233C.1
US-10-424-599-79977

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 17; Length 682;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 CTACCAAGCAGATCCTCTCCCGCCTGTTCGTTGATCCCAACCCCTC 231

RESULT 39
US-10-027-632-155300
; Sequence 155300, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155300
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-155300

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 13; Length 800;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 CTCTCCAAACCAGACCCCTTCCCTGGCCCTTGACATCCCAACCCCTC 296

RESULT 40
US-10-027-632-155300
; Sequence 155300, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155300
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-155300

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 13; Length 800;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 CTCTCCAAACCAGACCCCTTCCCTGGCCCTTGACATCCCAACCCCTC 296

RESULT 41
US-10-027-632-155023
; Sequence 155023, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155023
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-155023

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 13; Length 826;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCA 49
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 CCTCCCACTACGGTTCTGTGTCTCCCAACCCCTGCTCTCCCAACACTGCA 145

RESULT 42
US-10-027-632-155024
; Sequence 155024, Application US/10027632
; Publication No. US20020198371A1
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155300
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-155300

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 17; Length 800;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 CTCTCCAAACCAGACCCCTTCCCTGGCCCTTGACATCCCAACCCCTC 296

RESULT 41
US-10-027-632-155023
; Sequence 155023, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155023
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-155023

Query Match
Best Local Similarity 46.8%; Score 23.4; DB 13; Length 826;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCA 49
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 CCTCCCACTACGGTTCTGTGTCTCCCAACCCCTGCTCTCCCAACACTGCA 145

RESULT 42
US-10-027-632-155024
; Sequence 155024, Application US/10027632
; Publication No. US20020198371A1
```


Query Match	46.8%	Score 23.4;	DB 17;	Length 826;
Best Local Similarity	67.3%	Pred. No. 67;		

Query Match 46.8%; Score 23.4; DB 18; Length 38146;
Best Local Similarity 73.2%; Pred. No. 49;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 36876 ACACATCATCGCTGCTTTCCACCCCTGGATTTCACGCC 36836
|||||

RESULT 46

US-10-719-993-7051
; Sequence 7051, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7051
; LENGTH: 52128
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(52128)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-7051

Query Match 46.8%; Score 23.4; DB 18; Length 52128;
Best Local Similarity 73.2%; Pred. No. 48;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 ACCACCCCTGGATCCTTGCCCGCCCTTGAATTCACCAAC 44
|||||

Db 43890 ACACATCATCGCTGCTTTCCACCCCTGGATTTCACGCC 43930
|||||

RESULT 47

US-10-741-600-18002/c
; Sequence 18002, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18002
; LENGTH: 125560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(125560)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-18002

Query Match 46.8%; Score 23.4; DB 19; Length 125560;
Best Local Similarity 67.3%; Pred. No. 44;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 CTACCCACCTGGATCCTTGCCCGCCCTTGAATTCACCAACCTCCAC 50
|||||

Db 94581 CTCCTCAACCCAGACCCCTTTCCTGGCCCTTGACATCCCAACACCCCTC 94533
|||||

RESULT 48

US-10-425-115-93962
; Sequence 93962, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 93962
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_185680C.1
US-10-425-115-93962

Query Match 46.4%; Score 23.2; DB 18; Length 593;
Best Local Similarity 70.5%; Pred. No. 80;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCCGCCCTTGAATTCACCAAC 44
|||||

Db 393 CCTCTCTCTGGATCTTTTGGCCGCTCCTTGACTTCTCTGTCC 436
|||||

RESULT 49

US-10-437-963-10623
; Sequence 10623, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 10623
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_16926C.1
US-10-437-963-10623

Query Match 46.0%; Score 23; DB 18; Length 229;
Best Local Similarity 68.1%; Pred. No. 1e+02;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CCTACCCACCTGGATCCTTGCCCGCCCTTGAATTCACCAACCTC 47
|||||

Db 24 CCACCCCTCCCGGACCCCTTGATCTACCCATAAAATTCGCGCCGC 70
|||||

RESULT 50

US-10-767-701-8694
; Sequence 8694, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

```

; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8694
; LENGTH: 733
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS504_1
US-10-767-701-8694

Query Match      46.0%; Score 23; DB 18; Length 733;
Best Local Similarity 68.1%; Pred. No. 93;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      4 ACCACCCCTGGATCCTTGGCCCGCCCTTGAATTCCCAACCCCTCCAC 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      43 ACGGCACCCGGATCCCGCGCCACGCTTTAATTCCTCCCTCGAC 89

```

Search completed: April 25, 2005, 15:34:05
 Job time : 639.138 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:07:14 ; Search time 464.276 Seconds

(without alignments)
2296.080 Million cell updates/sec

Title: US-10-010-476-14

Perfect score: 22

Sequence: 1 ATCCTTGCCCGCCCTTGAAAT 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

GenEmbl.*

1: gb_bai.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ats.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	BD078130
2	22	100.0	30	6	BD078125
3	22	100.0	30	6	BD078126
4	22	100.0	50	6	BD078129
5	19	86.4	179537	2	AC013513
6	19	86.4	189756	9	AC092327
7	18.8	85.5	206000	10	AC102248
8	18.8	85.5	215671	10	AC113018
9	18.4	83.6	728	9	HS326066
10	18.4	83.6	37392	2	AC019617
11	18.4	83.6	61159	2	AC021911
12	18.4	83.6	91748	10	AF177767
13	18.4	83.6	162876	3	AC008212
14	18.4	83.6	163152	9	CNS057D1
15	18.4	83.6	167642	2	AC015945
16	18.4	83.6	169892	2	AC037483
17	18.4	83.6	181904	3	AC008211
18	18.4	83.6	203381	2	AC018444
19	18.4	83.6	213688	3	AE003755

c	20	18.4	83.6	217421	9	AC006211	9	AC006211	Homo sapi
c	21	18.4	83.6	218371	9	AF001025	9	AF001025	Homo sapi
c	22	18.4	83.6	300129	1	AB017309	1	AB017309	Desulfot
c	23	17.8	80.9	3172	10	BC068993	10	BC068993	Mus muscu
c	24	17.8	80.9	5980	3	CBRG37M09	3	CBRG37M09	Caenorhab
c	25	17.8	80.9	11076	4	AY427787	4	AY427787	Bos tauru
c	26	17.8	80.9	110000	2	AC146850_1	2	AC146850_1	Continuati
c	27	17.8	80.9	131427	9	HS37J18	9	HS37J18	Human DNA
c	28	17.8	80.9	135356	8	AF004769	8	AF004769	Oryza sat
c	29	17.8	80.9	138435	9	AC073468	9	AC073468	Homo sapi
c	30	17.8	80.9	157185	2	AC064798	2	AC064798	Homo sapi
c	31	17.8	80.9	158419	10	AC137901	10	AC137901	Mus muscu
c	32	17.8	80.9	165264	2	AC142254	2	AC142254	Mus muscu
c	33	17.8	80.9	184143	2	AC141993	2	AC141993	Rattus no
c	34	17.8	80.9	202267	9	DJ293M10	9	DJ293M10	Homo sapi
c	35	17.8	80.9	207930	2	AC120203	2	AC120203	Pan trogl
c	36	17.8	80.9	214493	2	AC097047	2	AC097047	Rattus no
c	37	17.8	80.9	226406	2	AC095975	2	AC095975	Rattus no
c	38	17.8	80.9	227369	2	AC114469	2	AC114469	Rattus no
c	39	17.8	80.9	227538	10	AC002397	10	AC002397	Mouse chr
c	40	17.8	80.9	234597	2	AC127638	2	AC127638	Rattus no
c	41	17.8	80.9	257605	2	AC120627	2	AC120627	Rattus no
c	42	17.8	80.9	263879	2	AC096060	2	AC096060	Rattus no
c	43	17.8	80.9	267829	2	AC120979	2	AC120979	Rattus no
c	44	17.8	80.9	267946	2	AC125633	2	AC125633	Rattus no
c	45	17.8	80.9	274467	2	AC113707	2	AC113707	Rattus no
c	46	17.8	80.9	336873	2	AC073666	2	AC073666	Mus muscu
c	47	17.4	79.1	180	6	CQ662027	6	CQ662027	Sequence
c	48	17.4	79.1	262	6	CQ697578	6	CQ697578	Sequence
c	49	17.4	79.1	277	6	CQ664542	6	CQ664542	Sequence
c	50	17.4	79.1	461	6	CQ398666	6	CQ398666	Sequence
c	51	17.4	79.1	528	11	BV209823	11	BV209823	Sequence
c	52	17.4	79.1	940	9	HSCHON03	9	HSCHON03	Mouse sapien
c	53	17.4	79.1	1708	9	BC036360	9	BC036360	Homo sapi
c	54	17.4	79.1	1739	9	BC073974	9	BC073974	Homo sapi
c	55	17.4	79.1	110000	1	BX571965_27	1	BX571965_27	Continuati
c	56	17.4	79.1	139471	2	AC139982	2	AC139982	Rattus no
c	57	17.4	79.1	173254	10	AL732396	10	AL732396	Mouse DNA
c	58	17.4	79.1	175262	9	AC021491	9	AC021491	Homo sapi
c	59	17.4	79.1	183674	2	AC133129	2	AC133129	Rattus no
c	60	17.4	79.1	197751	2	AC140075	2	AC140075	Rattus no
c	61	17.4	79.1	220873	2	AC120258	2	AC120258	Rattus no
c	62	17.4	79.1	223318	2	AC095389	2	AC095389	Rattus no
c	63	17.4	79.1	239416	2	AC130914	2	AC130914	Rattus no
c	64	17.4	79.1	248825	2	AC131555	2	AC131555	Rattus no
c	65	17.4	79.1	261688	2	AC111685	2	AC111685	Rattus no
c	66	17.4	79.1	307040	1	AE017315	1	AE017315	Desulfot
c	67	17.2	78.2	639	10	RNU34685	10	RNU34685	Rattus norv
c	68	17.2	78.2	84642	2	AC134360_3	2	AC134360_3	Continuati
c	69	17.2	78.2	92969	2	AC010844	2	AC010844	Drosophil
c	70	17.2	78.2	110000	2	AL929091_11	2	AL929091_11	Continuati
c	71	17.2	78.2	121081	3	AC024214	3	AC024214	Caenorhab
c	72	17.2	78.2	121081	3	AC024873	3	AC024873	Caenorhab
c	73	17.2	78.2	141806	2	AC006908	2	AC006908	Caenorhab
c	74	17.2	78.2	143786	3	AC024859	3	AC024859	Caenorhab
c	75	17.2	78.2	148894	8	AC121489	8	AC121489	Oryza sat
c	76	17.2	78.2	163138	10	AC135081	10	AC135081	Mus muscu
c	77	17.2	78.2	166977	9	AC007198	9	AC007198	Homo sapi
c	78	17.2	78.2	167389	9	AL772366	9	AL772366	Mouse DNA
c	79	17.2	78.2	169218	9	AL355297	9	AL355297	Human DNA
c	80	17.2	78.2	170945	2	AC079911	2	AC079911	Mus muscu
c	81	17.2	78.2	175838	2	AC116356	2	AC116356	Homo sapi
c	82	17.2	78.2	181466	9	AL359979	9	AL359979	Human DNA
c	83	17.2	78.2	201933	2	AC141729	2	AC141729	Apis mell
c	84	17.2	78.2	210291	10	AC129593	10	AC129593	Mus muscu
c	85	17.2	78.2	210306	10	AC141899	10	AC141899	Mus muscu
c	86	17.2	78.2	210636	9	AC006443	9	AC006443	Homo sapi
c	87	17.2	78.2	213593	2	AC131402	2	AC131402	Rattus no
c	88	17.2	78.2	214604	10	AC104898	10	AC104898	Mus muscu

c 93	17.2	78.2	218557	2	AC125811	Rattus no	166	16.8	76.4	98968	10	AL929498	AL929498 Mouse DNA
c 94	17.2	78.2	220626	2	AC134739	Rattus no	167	16.8	76.4	102389	3	AC004546	AC004546 Drosophila
c 95	17.2	78.2	233713	2	AC098341	Rattus no	168	16.8	76.4	104834	9	AL355178	AL355178 Human DNA
c 96	17.2	78.2	233849	2	AC113745	Rattus no	c 159	16.8	76.4	110000	1	AE017180_14	Continuation (15 o
c 97	17.2	78.2	238566	2	AC096795	Rattus no	c 170	16.8	76.4	110000	2	AC152779_2	Continuation (3 of
c 98	17.2	78.2	241243	2	AC094206	Rattus no	c 171	16.8	76.4	110702	9	AC097470	AC097470 Homo sapi
c 99	17.2	78.2	242321	2	AC095981	Rattus no	c 172	16.8	76.4	136799	8	AP004564	AP004564 Oryza sat
c 100	17.2	78.2	243639	2	AC096223	Rattus no	c 173	16.8	76.4	140414	2	AC136548	AC136548 Rattus no
c 101	17.2	78.2	243639	2	AC096223	Rattus no	c 174	16.8	76.4	141675	2	AP000405	AP000405 Homo sapi
c 102	17.2	78.2	245333	2	AC106184	Rattus no	c 175	16.8	76.4	141781	10	AC124327	AC124327 Mus muscu
c 103	17.2	78.2	245452	2	AC094820	Rattus no	c 176	16.8	76.4	148507	9	AC008784	AC008784 Homo sapi
c 104	17.2	78.2	246668	2	AC096521	Rattus no	c 177	16.8	76.4	151037	9	AC092620	AC092620 Homo sapi
c 105	17.2	78.2	249762	2	AC111422	Rattus no	c 178	16.8	76.4	151405	2	AC011172	AC011172 Homo sapi
c 106	17.2	78.2	251086	2	AC127139	Rattus no	c 179	16.8	76.4	152327	2	AC060759	AC060759 Homo sapi
c 107	17.2	78.2	255725	2	AC134804	Rattus no	c 180	16.8	76.4	153248	2	AC112500	AC112500 Homo sapi
c 108	17.2	78.2	272347	2	AC106694	Rattus no	c 181	16.8	76.4	153693	10	AC127553	AC127553 Mus muscu
c 109	17.2	78.2	295763	2	AC109373	Rattus no	c 182	16.8	76.4	162700	2	AC024105	AC024105 Homo sapi
c 110	17.2	78.2	298960	2	AC006896	Caenorhab	c 183	16.8	76.4	162700	2	AC024105	AC024105 Homo sapi
c 111	17.2	78.2	299015	2	AC006842	Caenorhab	c 184	16.8	76.4	163283	2	AC112980	AC112980 Mus muscu
c 112	17.2	78.2	299015	2	AC006842	Caenorhab	c 185	16.8	76.4	165697	3	AC092222	AC092222 Drosophila
c 113	17.2	78.2	313602	8	AP006168	Oryza sat	c 186	16.8	76.4	165697	3	AC092222	AC092222 Drosophila
c 114	17.2	77.3	136602	10	AK122261	AK122261 Mus muscu	c 187	16.8	76.4	167672	2	AC132806	AC132806 Homo sapi
c 115	17.2	77.3	138366	10	AC111733	AC111733 Rattus no	c 188	16.8	76.4	167700	2	AC079735	AC079735 Mus muscu
c 116	17.2	77.3	144581	2	AP004889	AP004889 Oryza sat	c 189	16.8	76.4	167700	2	AC079735	AC079735 Mus muscu
c 117	17.2	77.3	144581	2	AP004889	AP004889 Oryza sat	c 190	16.8	76.4	168125	9	AC016044	AC016044 Homo sapi
c 118	17.2	77.3	182990	2	AC131453	AC131453 Strongylo	c 191	16.8	76.4	168125	9	AC016044	AC016044 Homo sapi
c 119	17.2	77.3	204413	10	AL645910	AL645910 Mouse DNA	c 192	16.8	76.4	168413	2	AC009877	AC009877 Homo sapi
c 120	17.2	77.3	229494	10	AL627328	AL627328 Mouse DNA	c 193	16.8	76.4	169117	9	AC087894	AC087894 Homo sapi
c 121	17.2	77.3	238671	2	AC125296	AC125296 Rattus no	c 194	16.8	76.4	169682	2	BS572636	BS572636 Drosophila
c 122	17.2	77.3	242582	2	AC128397	AC128397 Rattus no	c 195	16.8	76.4	171612	3	AC009461	AC009461 Drosophila
c 123	17.2	77.3	271451	2	AP008503	AP008503 Rattus no	c 196	16.8	76.4	171612	3	AC009461	AC009461 Drosophila
c 124	17.2	77.3	271451	2	AP008503	AP008503 Rattus no	c 197	16.8	76.4	172224	2	AC116813	AC116813 Mus muscu
c 125	17.2	77.3	298900	1	AP005937	AP005937 Brachyrih	c 198	16.8	76.4	172224	2	AC116813	AC116813 Mus muscu
c 126	17.2	77.3	298900	1	AP005937	AP005937 Brachyrih	c 199	16.8	76.4	174125	2	AC110037	AC110037 Mus muscu
c 127	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 200	16.8	76.4	174125	2	AC110037	AC110037 Mus muscu
c 128	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 201	16.8	76.4	175878	2	AC122554	AC122554 Mus muscu
c 129	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 202	16.8	76.4	175878	2	AC122554	AC122554 Mus muscu
c 130	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 203	16.8	76.4	176708	9	AC039056	AC039056 Homo sapi
c 131	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 204	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 132	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 205	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 133	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 206	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 134	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 207	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 135	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 208	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 136	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 209	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 137	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 210	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 138	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 211	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 139	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 212	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 140	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 213	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 141	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 214	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 142	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 215	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 143	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 216	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 144	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 217	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 145	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 218	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 146	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 219	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 147	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 220	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 148	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 221	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 149	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 222	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 150	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 223	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 151	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 224	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 152	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 225	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 153	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 226	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 154	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 227	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 155	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 228	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 156	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 229	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 157	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 230	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 158	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 231	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 159	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 232	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 160	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 233	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 161	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 234	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 162	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 235	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 163	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 236	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 164	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 237	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi
c 165	17.2	77.3	347	6	CQ442367	CQ442367 Sequence	c 238	16.8	76.4	177076	9	AP000487	AP000487 Homo sapi

239	16.8	76.4	234379	2	AC098908	AC098908 Rattus no	c 312	16.4	74.5	174969	2	AC151745	AC151745 Mus muscu
240	16.8	76.4	242596	2	AC134759	AC134759 Rattus no	313	16.4	74.5	175286	2	AC046197	AC046197 Homo sapi
241	16.8	76.4	243432	2	AC125968	AC125968 Rattus no	314	16.4	74.5	176735	3	AC007810	AC007810 Drosophi
242	16.8	76.4	244146	2	AC126704	AC126704 Rattus no	315	16.4	74.5	180727	3	AC140278	AC140278 Mus muscu
243	16.8	76.4	244389	2	AC111858	AC111858 Rattus no	316	16.4	74.5	181927	3	AC010573	AC010573 Drosophi
244	16.8	76.4	245878	2	AC125842	AC125842 Rattus no	317	16.4	74.5	183046	2	AC140775	AC140775 Mus muscu
245	16.8	76.4	246011	10	AL929585	AL929585 Mouse DNA	c 318	16.4	74.5	183190	9	AC093327	AC093327 Homo sapi
246	16.8	76.4	252968	2	AC112300	AC112300 Rattus no	c 319	16.4	74.5	185825	10	AC087114	AC087114 Mus muscu
247	16.8	76.4	253101	3	AE003718	AE003718 Drosophi	320	16.4	74.5	186860	9	AC026471	AC026471 Homo sapi
248	16.8	76.4	255345	1	AC114514	AC114514 Rattus no	321	16.4	74.5	188062	9	AC007385	AC007385 Homo sapi
249	16.8	76.4	257860	2	AC127719	AC127719 Rattus no	c 322	16.4	74.5	189622	10	AL589681	AL589681 Mouse DNA
250	16.8	76.4	264814	2	AC122631	AC122631 Rattus no	c 323	16.4	74.5	190194	9	AC068025	AC068025 Homo sapi
251	16.8	76.4	266938	2	AC121426	AC121426 Rattus no	c 324	16.4	74.5	191118	2	AC148890	AC148890 Ocolemur
252	16.8	76.4	276289	1	AE017306	AE017306 Thermus t	325	16.4	74.5	191426	2	BX927352	BX927352 Panio rer
253	16.8	76.4	291417	1	AC012825	AC012825 Rattus t	c 326	16.4	74.5	194742	9	AC092047	AC092047 Homo sapi
254	16.8	76.4	295916	2	AC149224	AC149224 Mus muscu	c 327	16.4	74.5	194742	9	AC099054	AC099054 Homo sapi
255	16.8	76.4	300425	1	AP005041	AP005041 Streptomy	c 328	16.4	74.5	202418	2	AC150815	AC150815 Callithri
256	16.8	76.4	300933	1	AE016791	AE016791 Pseudomon	c 329	16.4	74.5	202563	9	AC021192	AC021192 Homo sapi
257	16.8	76.4	308447	3	AE003588	AE003588 Drosophi	330	16.4	74.5	203230	10	AL589670	AL589670 Mouse DNA
258	16.8	76.4	334222	3	AE003445	AE003445 Drosophi	331	16.4	74.5	203581	10	AL451076	AL451076 Mouse DNA
259	16.8	76.4	346294	1	AP002999	AP002999 Mesorhizo	332	16.4	74.5	207694	9	AL354384	AL354384 Human DNA
260	16.8	74.5	330	6	AX452112	AX452112 Sequence	333	16.4	74.5	215130	5	AL954320	AL954320 Zebrafish
261	16.4	74.5	1250	3	AY118600	AY118600 Drosophi	c 334	16.4	74.5	215647	2	AC068591	AC068591 Homo sapi
262	16.4	74.5	1256	6	CQ613917	CQ613917 Sequence	c 335	16.4	74.5	215647	2	AC026386	AC026386 Mus muscu
263	16.4	74.5	1749	10	AX388959	AX388959 Mus muscu	c 336	16.4	74.5	219316	2	AC111278	AC111278 Rattus no
264	16.4	74.5	1783	10	BC026774	BC026774 Mus muscu	c 337	16.4	74.5	219566	2	AC102664	AC102664 Mus muscu
265	16.4	74.5	2917	9	BC054514	BC054514 Homo sapi	338	16.4	74.5	223991	10	AL453300	AL453300 Mouse DNA
266	16.4	74.5	3549	6	CQ613916	CQ613916 Sequence	339	16.4	74.5	224818	10	AL845300	AL845300 Mouse DNA
267	16.4	74.5	4739	6	AR225786	AR225786 Sequence	340	16.4	74.5	227555	10	AC108416	AC108416 Mus muscu
268	16.4	74.5	7786	9	AF307340	AF307340 Homo sapi	c 341	16.4	74.5	228051	2	AC097434	AC097434 Rattus no
269	16.4	74.5	32503	6	CQ588929	CQ588929 Sequence	342	16.4	74.5	232278	2	AC087876	AC087876 Mus muscu
270	16.4	74.5	37947	8	SCU39205	SCU39205 Saccharomyc	343	16.4	74.5	233305	2	AC025751	AC025751 Mus muscu
271	16.4	74.5	56081	9	AC068588	AC068588 Homo sapi	344	16.4	74.5	236675	2	BX901904	BX901904 Danio rer
272	16.4	74.5	57085	2	AC017905	AC017905 Drosophi	c 345	16.4	74.5	237719	2	AC112284	AC112284 Rattus no
273	16.4	74.5	61412	2	AC021493	AC021493 Homo sapi	346	16.4	74.5	238472	9	AC010290	AC010290 Homo sapi
274	16.4	74.5	62763	2	AC069478	AC069478 Homo sapi	c 347	16.4	74.5	239162	5	BX927077	BX927077 Zebrafish
275	16.4	74.5	72932	2	AC015300	AC015300 Drosophi	c 348	16.4	74.5	240151	10	AC098877	AC098877 Mus muscu
276	16.4	74.5	75354	5	BX308397	Continuation (4 of	c 349	16.4	74.5	241228	2	AC103429	AC103429 Rattus no
277	16.4	74.5	87359	5	BX308397	Continuation (5 of	c 350	16.4	74.5	241783	2	AC106100	AC106100 Rattus no
278	16.4	74.5	90256	9	AP001582	AP001582 Homo sapi	c 351	16.4	74.5	242203	2	AC099230	AC099230 Rattus no
279	16.4	74.5	90320	9	AP003326	AP003326 Homo sapi	352	16.4	74.5	242324	10	AC093478	AC093478 Mus muscu
280	16.4	74.5	91034	2	AC025911	AC025911 Mus muscu	353	16.4	74.5	244551	2	AC119584	AC119584 Rattus no
281	16.4	74.5	110000	1	AE000516	Continuation (2 of	c 354	16.4	74.5	245669	2	AC095299	AC095299 Rattus no
282	16.4	74.5	110000	4	AC151843	Continuation (5 of	c 355	16.4	74.5	246805	2	AC098750	AC098750 Rattus no
283	16.4	74.5	110000	3	AC151846	Continuation (4 of	356	16.4	74.5	246962	2	AC073811	AC073811 Mus muscu
284	16.4	74.5	111951	9	AC010169	AC010169 Homo sapi	c 357	16.4	74.5	247619	2	AC107466	AC107466 Rattus no
285	16.4	74.5	121337	2	AC146934	AC146934 Sus scrofa	c 358	16.4	74.5	250046	2	AC113592	AC113592 Mus muscu
286	16.4	74.5	126738	2	AC148199	AC148199 Ornithorh	c 359	16.4	74.5	251452	2	AC097208	AC097208 Rattus no
287	16.4	74.5	128136	1	AF440524	AF440524 Pseudomon	c 360	16.4	74.5	253261	2	AC109368	AC109368 Rattus no
288	16.4	74.5	132671	10	AL732587	AL732587 Mouse DNA	c 361	16.4	74.5	260347	2	AC111702	AC111702 Rattus no
289	16.4	74.5	139280	9	AC099052	AC099052 Homo sapi	c 362	16.4	74.5	262798	2	AC111702	AC111702 Rattus no
290	16.4	74.5	142013	2	BX957298	BX957298 Homo rer	363	16.4	74.5	265861	2	AC069274	AC069274 Mus muscu
291	16.4	74.5	146500	9	AC004453	AC004453 Homo sapi	364	16.4	74.5	270351	2	AC106615	AC106615 Rattus no
292	16.4	74.5	148640	5	AF056116	AF056116 Fugu rubr	c 365	16.4	74.5	270351	2	AC106615	AC106615 Rattus no
293	16.4	74.5	154235	10	AC123029	AC123029 Mus muscu	c 366	16.4	74.5	270456	2	AC111575	AC111575 Rattus no
294	16.4	74.5	154353	2	AC149503	AC149503 Xenopus t	367	16.4	74.5	278708	3	AE003535	AE003535 Drosophi
295	16.4	74.5	159752	2	AL114241	AL114241 Rattus no	c 368	16.4	74.5	282407	2	AC111411	AC111411 Rattus no
296	16.4	74.5	159771	9	AL139142	AL139142 Human DNA	c 369	16.4	74.5	282861	2	AC111701	AC111701 Rattus no
297	16.4	74.5	160237	5	BX248397	BX248397 Zebrafish	c 370	16.4	74.5	283363	2	AC095687	AC095687 Rattus no
298	16.4	74.5	160366	3	AC069457	AC069457 Drosophi	c 371	16.4	74.5	341957	1	BX842572	BX842572 Mycobacte
299	16.4	74.5	162163	2	AC021635	AC021635 Homo sapi	c 372	16.4	74.5	343050	1	BX248334	BX248334 Mycobacte
300	16.4	74.5	163788	5	BX571822	BX571822 Zebrafish	c 373	16.4	74.5	347965	2	AC135411	AC135411 Rattus no
301	16.4	74.5	165294	4	AC140350	AC140350 Mus muscu	c 374	16.2	73.6	222	6	AX439734	AX439734 Sequence
302	16.4	74.5	165737	10	AC138025	AC138025 Mus muscu	375	16.2	73.6	359	6	CQ657994	CQ657994 Sequence
303	16.4	74.5	166001	9	HS4462018	AL132765 Human DNA	376	16.2	73.6	482	6	CQ481698	CQ481698 Sequence
304	16.4	74.5	166384	2	AC148889	AC148889 Ocolemur	377	16.2	73.6	486	11	G96504	G96504 S208P6508RG
305	16.4	74.5	166421	10	AC123532	AC123532 Mus muscu	c 378	16.2	73.6	502	6	AX971677	AX971677 Sequence
306	16.4	74.5	168317	2	AC048329	AC048329 Homo sapi	c 379	16.2	73.6	502	6	AX971677	AX971677 Sequence
307	16.4	74.5	169424	2	AL929203	AL929203 Homo sapi	c 380	16.2	73.6	502	6	BD110396	BD110396 EST and e
308	16.4	74.5	170848	2	AC093630	AC093630 Homo sapi	c 381	16.2	73.6	506	6	BD112977	BD112977 Sequence
309	16.4	74.5	171115	9	AC068385	AC068385 Homo sapi	c 382	16.2	73.6	506	6	AX969811	AX969811 Sequence
310	16.4	74.5	171197	2	AC149510	AC149510 Xenopus t	c 383	16.2	73.6	506	6	BD108530	BD108530 EST and e
311	16.4	74.5	174187	2	AP002792	AP002792 Homo sapi	384	16.2	73.6	643	8	AF401143	AF401143 Acer kung

385	16.2	73.6	697	8	AF020384	AF020384 Acer plat	C 458	16.2	73.6	53336	9	AC092600	AC092600 Homo sapi
386	16.2	73.6	704	6	AX070909	AX070909 Sequence	C 459	16.2	73.6	53424	3	AB055861	AB055861 Procambar
387	16.2	73.6	980	6	CQ722768	CQ722768 Sequence	C 460	16.2	73.6	54475	9	AL136122	AL136122 Human DNA
388	16.2	73.6	980	6	CQ723944	CQ723944 Sequence	C 461	16.2	73.6	54911	9	BX248307	BX248307 Human DNA
389	16.2	73.6	992	6	CQ722767	CQ722767 Sequence	C 462	16.2	73.6	58554	2	AC100623	AC100623 Mus muscu
390	16.2	73.6	992	6	CQ723943	CQ723943 Sequence	C 463	16.2	73.6	62145	2	AC116114	AC116114 Mus muscu
391	16.2	73.6	1025	9	BC000691	BC000691 Homo sapi	C 464	16.2	73.6	63771	9	AL590425	AL590425 Human DNA
392	16.2	73.6	1125	14	EC0095014	U95014 Ecotrops o	C 465	16.2	73.6	66980	2	AC013796	AC013796 Homo sapi
393	16.2	73.6	1128	6	AX658070	AX658070 Sequence	C 466	16.2	73.6	70000	6	AR215862	AR215862 Sequence
394	16.2	73.6	1155	6	BD268272	BD268272 Novel odo	C 467	16.2	73.6	71503	2	AC090205	AC090205 Homo sapi
395	16.2	73.6	1155	6	CQ582546	CQ582546 Sequence	C 468	16.2	73.6	72240	9	CR759873	CR759873 Human DNA
396	16.2	73.6	1155	6	AR382440	AR382440 Sequence	C 469	16.2	73.6	73317	9	AY522921	AY522921 Homo sapi
397	16.2	73.6	1354	6	BD268321	BD268321 Novel odo	C 470	16.2	73.6	73679	9	HSDJ49503	AL121587 Human DNA
398	16.2	73.6	1354	6	AR382489	AR382489 Sequence	C 471	16.2	73.6	75074	9	CR788240	CR788240 Human DNA
399	16.2	73.6	1428	1	AF306346	AF306346 Mycoplasma	C 472	16.2	73.6	75313	2	AC017682	AC017682 Drosophil
400	16.2	73.6	1543	9	AK097618	AK097618 Homo sapi	C 473	16.2	73.6	75337	2	AC101577	AC101577 Mus muscu
401	16.2	73.6	1743	10	BC012700	BC012700 Mus muscu	C 474	16.2	73.6	76734	2	AC023218	AC023218 Homo sapi
402	16.2	73.6	1787	6	AX748149	AX748149 Sequence	C 475	16.2	73.6	77355	2	AC139018	AC139018 Homo sapi
403	16.2	73.6	1787	9	AK093473	AK093473 Homo sapi	C 476	16.2	73.6	77463	2	AC019795	AC019795 Drosophil
404	16.2	73.6	2073	9	BC016843	BC016843 Homo sapi	C 477	16.2	73.6	77691	9	AC120118	AC120118 Homo sapi
405	16.2	73.6	2149	6	AX834741	AX834741 Sequence	C 478	16.2	73.6	79568	2	AC151725	AC151725 Medicago
406	16.2	73.6	2149	6	AK097388	AK097388 Homo sapi	C 479	16.2	73.6	81482	2	AL1391685	AL1391685 Homo sapi
407	16.2	73.6	2223	6	AX748139	AX748139 Sequence	C 480	16.2	73.6	85411	9	AC079782	AC079782 Homo sapi
408	16.2	73.6	2223	9	AK093458	AK093458 Homo sapi	C 481	16.2	73.6	88017	9	AC128655	AC128655 Homo sapi
409	16.2	73.6	2233	9	BC063449	BC063449 Homo sapi	C 482	16.2	73.6	90104	6	CQ588086	CQ588086 Sequence
410	16.2	73.6	2248	8	AK105640	AK105640 Oryza sat	C 483	16.2	73.6	90244	9	AB023051	AB023051 Homo sapi
411	16.2	73.6	2290	9	BC011799	BC011799 Homo sapi	C 484	16.2	73.6	93670	9	AP003502	AP003502 Homo sapi
412	16.2	73.6	2298	6	AX834251	AX834251 Sequence	C 485	16.2	73.6	94514	9	AL953874	AL953874 Human DNA
413	16.2	73.6	2582	1	NGPR10FA	AK096591 Homo sapi	C 486	16.2	73.6	95375	9	HS228A9	AL022322 Human DNA
414	16.2	73.6	2582	6	AX675037	Z25461 N.gonorrhoe	C 487	16.2	73.6	96512	9	AC022539	AC022539 Homo sapi
415	16.2	73.6	2622	6	AX675037	AX675037 Sequence	C 488	16.2	73.6	100521	9	AC119397	AC119397 Homo sapi
416	16.2	73.6	2634	6	AX675035	AX675035 Sequence	C 489	16.2	73.6	101391	2	AC134153	AC134153 Rattus no
417	16.2	73.6	3153	8	AK107241	AK107241 Oryza sat	C 490	16.2	73.6	102581	9	AL139390	AL139390 Human DNA
418	16.2	73.6	3234	6	CQ582545	CQ582545 Sequence	C 491	16.2	73.6	110000	1	AE017261	Continuation (3 of
419	16.2	73.6	3243	6	CQ718190	CQ718190 Sequence	C 492	16.2	73.6	110000	1	AE017282	Continuation (19 o
420	16.2	73.6	5405	10	HMA73199	U73199 Mus musculu	C 493	16.2	73.6	110000	1	CR522870	Continuation (28 o
421	16.2	73.6	6013	9	NSA318892	AJ318892 Homo sapi	C 494	16.2	73.6	110000	2	AC091359	Rattus no
422	16.2	73.6	6852	6	AX251948	AX251948 Sequence	C 495	16.2	73.6	110000	2	AC094316	Continuation (3 of
423	16.2	73.6	6852	6	AX348743	AX348743 Sequence	C 496	16.2	73.6	110000	2	AC095155	Continuation (2 of
424	16.2	73.6	6968	6	CQ726039	CQ726039 Sequence	C 497	16.2	73.6	110000	2	AC095467	Continuation (3 of
425	16.2	73.6	7008	6	AX741281	AX741281 Sequence	C 498	16.2	73.6	110000	2	AC097152	Rattus no
426	16.2	73.6	7008	6	AB027567	AB027567 Homo sapi	C 499	16.2	73.6	110000	2	AC097152	Continuation (2 of
427	16.2	73.6	8715	14	AF132668	AF204680 Dunaliell	C 500	16.2	73.6	110000	2	AC108345	Rattus no
428	16.2	73.6	9489	8	AF204680	AF204680 Dunaliell							
429	16.2	73.6	10029	1	AE013355	AE013355 Methanosa							
430	16.2	73.6	10354	1	AE009305	AE009305 Agrobacte							
431	16.2	73.6	11229	1	AE000804	AE000804 Methanoba							
432	16.2	73.6	11573	1	AE000809	AE000809 Agrobacte							
433	16.2	73.6	11852	14	CICIV12	Z58828 Chelonius in							
434	16.2	73.6	13135	1	AE014501	AE014501 Brucella							
435	16.2	73.6	14375	1	AE009631	AE009631 Brucella							
436	16.2	73.6	15218	2	AC014602	AC014602 Drosophil							
437	16.2	73.6	1528	14	AF107100	AF107100 Ecotrops							
438	16.2	73.6	16076	3	AB055862	AB055862 Procambar							
439	16.2	73.6	16567	5	AP004429	AP004429 Monocentr							
440	16.2	73.6	16751	5	AP004440	AP004440 Dactylopt							
441	16.2	73.6	18234	6	AX151455	AX151455 Sequence							
442	16.2	73.6	18234	6	AX658052	AX658052 Sequence							
443	16.2	73.6	20137	6	AR148845	AR148845 Sequence							
444	16.2	73.6	20138	6	AR148652	AR148652 Sequence							
445	16.2	73.6	23071	6	AR148847	AR148847 Sequence							
446	16.2	73.6	28080	4	SSCR39YR1	Z49778 S. scrofa ge							
447	16.2	73.6	32816	6	AX571959	AX571959 Sequence							
448	16.2	73.6	36216	9	BX327283	BX327283 Human DNA							
449	16.2	73.6	36690	2	AC014124	AC014124 Drosophil							
450	16.2	73.6	37349	9	AC010524	AC010524 Homo sapi							
451	16.2	73.6	38491	9	AC006049	AC006049 Homo sapi							
452	16.2	73.6	39749	3	DMC4984	AL031128 Drosophil							
453	16.2	73.6	41381	2	AC149383	AC149383 Phakopsor							
454	16.2	73.6	42715	3	CBRG45M20	AC084648 Caenorhab							
455	16.2	73.6	42913	2	AC090704	AC090704 Homo sapi							
456	16.2	73.6	44118	9	AC006165	AC006165 Homo sapi							
457	16.2	73.6	50777	9	AC008535	AC008535 Homo sapi							

ALIGNMENTS

RESULT 1	BD078130	22 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD078130	Modulator of DNA cytosine-5 methyltransferase and method of using			
DEFINITION	the same.				
ACCESSION	BD078130				
VERSION	BD078130.1	GI:22623733			
KEYWORDS	JP 2001514862-A/14.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 22)				
AUTHORS	Reich, N. O. and Flynn, J.				
TITLE	Modulator of DNA cytosine-5 methyltransferase and method of using				
JOURNAL	Patent: JP 2001514862-A 14 18-SEP-2001;				
COMMENT	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA				
	OS Unidentified				
	PN JP 2001514862-A/14				
	PD 18-SEP-2001				
	PF 12-JUN-1998 JP 2000508978				
	PR 29-AUG-1997 US 60/057411				
	PI NORBERT O. REICH, JAMES FLYNN				
	PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00				
	CC Strandedness: Double;				


```
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..22
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..22
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCCGCCCTTGAAT 22
DB 1 ATCCTTGCCCCGCCCTTGAAT 22
RESULT 2
BD078125/c
LOCUS
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
BD078125 30 bp DNA linear PAT 27-AUG-2002
ACCESSION BD078125
VERSION BD078125.1 GI:22623728
KEYWORDS JP 2001514862-A/9.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 9 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/9
PD 18-SEP-2001
PR 12-JUN-1998 JP 2000508978
PF 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..30
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCCGCCCTTGAAT 22
DB 26 ATCCTTGCCCCGCCCTTGAAT 5
RESULT 3
BD078126
LOCUS
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
BD078126 30 bp DNA linear PAT 27-AUG-2002
ACCESSION BD078126
VERSION BD078126.1 GI:22623729
KEYWORDS JP 2001514862-A/10.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 10 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/10
PD 18-SEP-2001
PR 12-JUN-1998 JP 2000508978
PF 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..30
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
```

```
the same.
BD078126
LOCUS
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
BD078126 50 bp DNA linear PAT 27-AUG-2002
ACCESSION BD078126
VERSION BD078126.1 GI:22623732
KEYWORDS JP 2001514862-A/13.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 50)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 13 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/13
PD 18-SEP-2001
PR 12-JUN-1998 JP 2000508978
PF 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..50
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCCGCCCTTGAAT 22
DB 5 ATCCTTGCCCCGCCCTTGAAT 26
RESULT 4
BD078129
LOCUS
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
BD078129 50 bp DNA linear PAT 27-AUG-2002
ACCESSION BD078129
VERSION BD078129.1 GI:22623732
KEYWORDS JP 2001514862-A/13.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 50)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 13 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/13
PD 18-SEP-2001
PR 12-JUN-1998 JP 2000508978
PF 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..50
FT /organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
```

```

FEATURES
  source
    Location/Qualifiers
      1..50
        /organism="Unidentified".
        /organism="unidentified"
        /mol_type="genomic DNA"
        /db_xref="taxon:32644"
    ORIGIN
      Query Match
      Best Local Similarity 100.0%; Score 22; DB 6; Length 50;
      Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ATCCTTGCCCCCGCCCTTGAAT 22
    |||||
Db 15 ATCCTTGCCCCCGCCCTTGAAT 36

RESULT 5
AC013513 179537 bp DNA linear HTG 09-SEP-2000
LOCUS Homo sapiens clone RP11-115J15, WORKING DRAFT SEQUENCE, 21
DEFINITION
  unordered pieces.
ACCESSION AC013513
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 179537)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
  Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Bouckghalter,B.,
  Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
  Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
  Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
  Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
  Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
  Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
  McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
  Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
  Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
  Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
  Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
  Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141 USA
  On Sep 9, 2000 this sequence version replaced gi:6573938.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L1886
  Center clone name: 115_J_15
  ----- Summary Statistics
  Sequencing vector: M13; M7815; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.960731
  Consensus quality: 143563 bases at least Q40
  Consensus quality: 163588 bases at least Q30
  Consensus quality: 171970 bases at least Q20
  Insert size: 157000; agarose-fp
  Insert size: 177537; sum-of-contigs
  Quality coverage: 5.0 in Q20 bases; agarose-fp
  -----
  Location/Qualifiers
    1..2197
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="RP11-115J15"
      /clone_lib="RPC1-11 Human Male BAC"
    1..2197
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left"
    2298..3454
      /note="assembly_fragment"
    3555..5529
      /note="assembly_fragment"
    5630..7014
      /note="assembly_fragment"
    7115..11246
      /note="assembly_fragment"
    11347..14068
      /note="assembly_fragment"
    14169..18157
      /note="assembly_fragment"
  FEATURES
    source
      1..179537
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /clone="RP11-115J15"
        /clone_lib="RPC1-11 Human Male BAC"
      1..2197
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left"
      2298..3454
        /note="assembly_fragment"
      3555..5529
        /note="assembly_fragment"
      5630..7014
        /note="assembly_fragment"
      7115..11246
        /note="assembly_fragment"
      11347..14068
        /note="assembly_fragment"
      14169..18157
        /note="assembly_fragment"
  Quality coverage: 4.4 in Q20 bases; sum-of-contigs
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 21 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 2197: contig of 2197 bp in length
  * 2198 2297: gap of 100 bp
  * 2298 3454: contig of 1157 bp in length
  * 3455 3554: gap of 100 bp
  * 3555 5529: contig of 1975 bp in length
  * 5530 5629: gap of 100 bp
  * 5630 7014: contig of 1385 bp in length
  * 7015 7114: gap of 100 bp
  * 7115 11246: contig of 4132 bp in length
  * 11247 11346: gap of 100 bp
  * 11347 14068: contig of 2722 bp in length
  * 14069 14168: gap of 100 bp
  * 14169 18157: contig of 3989 bp in length
  * 18158 18257: gap of 100 bp
  * 18258 23215: contig of 4958 bp in length
  * 23216 23315: gap of 100 bp
  * 23316 30517: contig of 7202 bp in length
  * 30518 30617: gap of 100 bp
  * 30618 34502: contig of 3885 bp in length
  * 34503 34602: gap of 100 bp
  * 34603 39263: contig of 4661 bp in length
  * 39264 44470: contig of 5107 bp in length
  * 44471 44570: gap of 100 bp
  * 44571 50561: contig of 5991 bp in length
  * 50562 50661: gap of 100 bp
  * 50662 61183: contig of 10522 bp in length
  * 61184 61283: gap of 100 bp
  * 61284 69075: contig of 7792 bp in length
  * 69076 69175: gap of 100 bp
  * 69176 79364: contig of 10189 bp in length
  * 79365 79464: gap of 100 bp
  * 79465 90636: contig of 11172 bp in length
  * 90637 90736: gap of 100 bp
  * 90737 103766: contig of 13030 bp in length
  * 103767 103866: gap of 100 bp
  * 103867 124110: contig of 20244 bp in length
  * 124111 124210: gap of 100 bp
  * 124211 154260: contig of 30050 bp in length
  * 154261 154360: gap of 100 bp
  * 154361 179537: contig of 25177 bp in length.
  Location/Qualifiers
    1..179537
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="RP11-115J15"
      /clone_lib="RPC1-11 Human Male BAC"
    1..2197
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left"
    2298..3454
      /note="assembly_fragment"
    3555..5529
      /note="assembly_fragment"
    5630..7014
      /note="assembly_fragment"
    7115..11246
      /note="assembly_fragment"
    11347..14068
      /note="assembly_fragment"
    14169..18157
      /note="assembly_fragment"

```

```

/note="assembly_fragment"
18258..23215
/note="assembly_fragment"
23316..30517
/note="assembly_fragment"
30618..34502
/note="assembly_fragment"
34603..39263
/note="assembly_fragment"
39364..44470
/note="assembly_fragment"
44571..50561
/note="assembly_fragment"
50662..51183
/note="assembly_fragment"
61284..69075
/note="assembly_fragment"
69176..79364
/note="assembly_fragment"
79465..90636
/note="assembly_fragment"
90737..103766
/note="assembly_fragment"
103867..124110
/note="assembly_fragment"
124211..154260
/note="assembly_fragment"
154361..179537
/note="assembly_fragment"
clone_end:T7
vector_side:right"

misc_feature
18258..23215
/note="assembly_fragment"
23316..30517
/note="assembly_fragment"
30618..34502
/note="assembly_fragment"
34603..39263
/note="assembly_fragment"
39364..44470
/note="assembly_fragment"
44571..50561
/note="assembly_fragment"
50662..51183
/note="assembly_fragment"
61284..69075
/note="assembly_fragment"
69176..79364
/note="assembly_fragment"
79465..90636
/note="assembly_fragment"
90737..103766
/note="assembly_fragment"
103867..124110
/note="assembly_fragment"
124211..154260
/note="assembly_fragment"
154361..179537
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match      86.4%; Score 19; DB 2; Length 179537;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCCGCCCTTG 19
Db 173437 ATCCTTGCCCCGCCCTTG 173455

RESULT 6
AC092327
LOCUS Homo sapiens chromosome 16 clone RP11-15813, complete sequence.
DEFINITION AC092327 AC073996
ACCESSION AC092327.4 GI:38502332
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189756)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189756)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 189756)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 189756)
AUTHORS Stanford Human Genome Center and Los Alamos National Laboratory.
CONSTRM DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Submitted (25-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

/note="assembly_fragment"
18258..23215
/note="assembly_fragment"
23316..30517
/note="assembly_fragment"
30618..34502
/note="assembly_fragment"
34603..39263
/note="assembly_fragment"
39364..44470
/note="assembly_fragment"
44571..50561
/note="assembly_fragment"
50662..51183
/note="assembly_fragment"
61284..69075
/note="assembly_fragment"
69176..79364
/note="assembly_fragment"
79465..90636
/note="assembly_fragment"
90737..103766
/note="assembly_fragment"
103867..124110
/note="assembly_fragment"
124211..154260
/note="assembly_fragment"
154361..179537
/note="assembly_fragment"
clone_end:T7
vector_side:right"

misc_feature
18258..23215
/note="assembly_fragment"
23316..30517
/note="assembly_fragment"
30618..34502
/note="assembly_fragment"
34603..39263
/note="assembly_fragment"
39364..44470
/note="assembly_fragment"
44571..50561
/note="assembly_fragment"
50662..51183
/note="assembly_fragment"
61284..69075
/note="assembly_fragment"
69176..79364
/note="assembly_fragment"
79465..90636
/note="assembly_fragment"
90737..103766
/note="assembly_fragment"
103867..124110
/note="assembly_fragment"
124211..154260
/note="assembly_fragment"
154361..179537
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match      86.4%; Score 19; DB 9; Length 189756;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCCGCCCTTG 19
Db 84169 ATCCTTGCCCCGCCCTTG 84187

RESULT 7
AC102248/c
LOCUS Mus musculus chromosome 18, clone RP24-15819, complete sequence.
DEFINITION AC102248
ACCESSION AC102248.19 GI:33859954
VERSION HTG.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 206000)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Mus musculus chromosome 18, clone RP24-15819
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 206000)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R.,
Jones, C., Kanat, A., Katag, A., Kells, C., LaRocque, K.,
Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 206000)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

REFERENCE
AUTHORS      4 (bases 1 to 206000)
TITLE        Birren,B., Nusbaum,C. and Lander,B.
JOURNAL      Direct Submission
SUBMITTED    Submitted (19-AUG-2003) Whitehead Institute/MIT Center for Genome
JOURNAL      Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS      5 (bases 1 to 206000)
TITLE        Birren,B., Nusbaum,C., Lander,B., Abouelleil,A., Allen,N.,
JOURNAL      Anderson,M., Arachchi,H.M., Barna,N., Baetien,V., Bloom,T.,
SUBMITTED    Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
JOURNAL      Collimore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
SUBMITTED    Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
JOURNAL      Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
SUBMITTED    Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
JOURNAL      Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
SUBMITTED    Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
JOURNAL      Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
SUBMITTED    Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
JOURNAL      Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
SUBMITTED    Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
JOURNAL      O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Piere,N.,
SUBMITTED    Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
JOURNAL      Roman,J., Schauer,S., Schuppback,R., Seaman,S., Severy,P., Smith,C.,
SUBMITTED    Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
JOURNAL      Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
SUBMITTED    Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
JOURNAL      Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
COMMENT      Direct Submission
SUBMITTED    Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
JOURNAL      Research, 320 Charles Street, Cambridge, MA 02141, USA
SUBMITTED    On Aug 19, 2003 this sequence version replaced gi:33589945.
JOURNAL      All repeats were identified using RepeatMasker:
SUBMITTED    Smit, A.F.A. & Green, P. (1996-1997)
JOURNAL      http://ftp.genome.washington.edu/RM/RepeatMasker.html
FEATURES
source       1..206000
             Location/Qualifiers
             /organism="Mus musculus"
             /mol_type="Genomic DNA"
             /db_xref="taxon:10090"
             /chromosome="18"
             /map="18"
             /clone="RP24-15819"
             /clone_lib="RP24 Male Mouse BAC"
             1..1705
             /note="wgs_end extension
             clone_end:SP6"
             1706..1709
             /note="clone boundary
             clone_end:SP6
             site:Mbol"
             complement(11831..11841)
             /note="single clone coverage"
             complement(22887..22896)
             /note="single clone coverage"
             41205..43566
             /rpt_family="Lx2"
             repeat_region
             43583..44199
             /rpt_family="L1M2"
             repeat_region
             complement(44209..44442)
             /rpt_family="URR1B"
             repeat_region
             44735..44761
             /rpt_family="(TAA)n"
             repeat_region
             44924..44951
             /rpt_family="(GGGGA)n"
             repeat_region
             45069..45449
             /rpt_family="RLTR14"
             repeat_region
             /rpt_family="Charlie?"
             46341..46361
             /rpt_family="AT rich"
             repeat_region
             complement(48065..48176)
             /rpt_family="RMR1B"
             repeat_region
             48311..48365
             /rpt_family="MLT1A0"
             repeat_region
             48514..48548
             /rpt_family="MLT1A0"
             repeat_region
             48549..48654
             /rpt_family="PB1D9"
             repeat_region
             48655..48668
             /rpt_family="MLT1A0"
             repeat_region
             48669..48731
             /rpt_family="(GA)n"
             repeat_region
             48732..48842
             /rpt_family="MLT1A0"
             repeat_region
             48918..49049
             /rpt_family="B1_MM"
             repeat_region
             49052..49076
             /rpt_family="(CAAA)n"
             repeat_region
             49141..49179
             /rpt_family="AT rich"
             repeat_region
             49843..50044
             /rpt_family="L1M4"
             repeat_region
             50055..50200
             /rpt_family="B1_MM"
             repeat_region
             50201..50248
             /rpt_family="A-rich"
             repeat_region
             complement(50419..50571)
             /rpt_family="Tigger5"
             repeat_region
             complement(50637..50967)
             /rpt_family="NTD"
             repeat_region
             51047..51516
             /rpt_family="L1M4"
             repeat_region
             51752..52006
             /rpt_family="B4A"
             repeat_region
             52067..52230
             /rpt_family="ID_B1"
             repeat_region
             52247..52280
             /rpt_family="(GGA)n"
             repeat_region
             complement(52369..52419)
             /rpt_family="ORR1A1"
             repeat_region
             complement(52454..54279)
             /rpt_family="L1_MM"
             repeat_region
             complement(54275..55593)
             /rpt_family="L1_MM"
             repeat_region
             55594..56221
             /rpt_family="L1"
             repeat_region
             complement(56220..56444)
             /rpt_family="L1"
             repeat_region
             complement(56445..56683)
             /rpt_family="ORR1A2"
             repeat_region
             57359..57485
             /rpt_family="PB1D10"
             repeat_region
             57487..57837
             /rpt_family="L1MA9"
             repeat_region
             57846..57969
             /rpt_family="ID3"
             repeat_region
             57987..58094
             /rpt_family="L1MA9"
             repeat_region
             58104..58301
             /rpt_family="B4"
             repeat_region
             complement(58302..58696)
             /rpt_family="MTA"
             repeat_region
             58697..58747
             /rpt_family="B4"
             repeat_region
             58774..58858
             /rpt_family="L1MA9"
             repeat_region
             58870..58976
             /rpt_family="L1MA9"
             repeat_region
             complement(59055..59554)
             /rpt_family="RLTR14"
             repeat_region

```

```

repeat_region 61058..61101
/rpt_family="(TA)n"
repeat_region 63284..63356
/rpt_family="BC1_MM"
repeat_region 63287..63418
/rpt_family="ID_B1"
repeat_region 63463..63647
/rpt_family="URR1B"
repeat_region 63648..63840
/rpt_family="CYR11_MM"

Query Match      85.5%; Score 18.8; DB 10; Length 206000;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTCCCGCCGCTTGAAT 22
Db 124732 ATCCTTCCCGCCGCTTGAAT 124711

RESULT 8
AC113018/c
LOCUS          215671 bp      DNA      linear      ROD 02-AUG-2003
DEFINITION    Mus musculus, clone RP23-183C18, complete sequence.
ACCESSION     AC113018
VERSION       HTG.
KEYWORDS      Mus musculus (house mouse)
SOURCE        Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 215671)
AUTHORS       Birren,B., Nussbaum,C. and Lander,E.
TITLE         Mus musculus, clone RP23-183C18
JOURNAL       Unpublished
AUTHORS        2 (bases 1 to 215671)
              Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
              Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
              Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
              Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
              Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
              Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
              Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
              Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
              Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
              MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
              McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
              Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
              Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
              Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
              Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
              Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
              Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
              Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
              Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
              Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 215671)
              Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
              Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
              Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
              Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
              Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
              Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
              Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
              Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
              Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
              Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
              MacDonald,P., Major,J., Manning,J., Mienga,V.,
              Murdoch,M., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
              O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
              Pollara,V., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J.,
              Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
              Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
              Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
              Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
              Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 215671)
              Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
              Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
              Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
              Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
              Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
              Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
              Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
              Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
              Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
              Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
              MacDonald,P., Major,J., Manning,J., Mienga,V.,
              Murdoch,M., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
              O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
              Pollara,V., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J.,
              Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
              Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
              Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
              Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
              Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-AUG-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 2, 2003 this sequence version replaced gi:32469529.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genomic Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23498
Center clone name: 183_C18
----- Location/Qualifiers
1..215671
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /clone="RP23-183C18"
  /clone_libs="RPC1-23 Female Mouse BAC"
1..12378
  /note="wgs end extension
clone end:SP6"
repeat_region 1..25
/rpt_family="(GGAA)n"
repeat_region 173..203
/rpt_family="AT_rich"
repeat_region 1203..1612
/rpt_family="Lx"
repeat_region 1613..1637
/rpt_family="AT_rich"
repeat_region 2794..2855
/rpt_family="RSINE1"
repeat_region 3892..3926
/rpt_family="(CAAAA)n"
complement(4010..4193)
/rpt_family="B2_Mm2"

```

```
repeat_region 4496..4520
                /rpt_family="TTTTTC)n"
repeat_region 4521..4549
                /rpt_family=" (CTTTG)n"
repeat_region 5362..5617
                /rpt_family="Lx9"
repeat_region 5883..6005
                /rpt_family="RMER3"
                /complement(6216..6517)
repeat_region /rpt_family="MYSERV"
                /complement(6611..6759)
repeat_region /rpt_family="MYSERV"
                /complement(7475..7479)
                /note="<30 qual SINGL region"
                /complement(7486..7493)
                /note="<30 qual SINGL region"
                /complement(7513..7517)
                /note="<30 qual SINGL region"
                /complement(7527..7532)
                /note="<30 qual SINGL region"
                /complement(8184..8573)
                /rpt_family="IAPLTR2_MM"
                /complement(8184..8573)
repeat_region 9067..9460
                /rpt_family="Lx9"
                /complement(9461..9496)
                /rpt_family=" (CA)n"
                /complement(9499..9525)
                /rpt_family=" (GA)n"
                /complement(9526..9665)
                /rpt_family="Lx9"
                /complement(9746..9885)
                /rpt_family="GA-rich"
                /complement(9894..9969)
                /rpt_family="G-rich"
                /complement(9989..10204)
                /rpt_family="MLT1A0"
                /complement(11923..11952)
                /rpt_family="AT-rich"
                /complement(12375..12384)
                /note="clone boundary"
                /clone_end:SP6
                /site:ECORI
repeat_region complement(12723..12920)
                /rpt_family="L2"
repeat_region complement(13173..13339)
                /rpt_family="Lx8"
                /complement(13336..13410)
                /rpt_family="Lx8"
                /complement(13657..13827)
                /rpt_family="LIMCa"
                /complement(13845..13899)
                /rpt_family="T-rich"
                /complement(13954..14494)
                /rpt_family="LIMCa"
                /complement(14807..14888)
                /rpt_family="A-rich"
                /complement(14900..15021)
                /rpt_family="MLT1E3"
                /complement(15222..15345)
                /rpt_family=" (CA)n"
                /complement(15355..15662)
                /rpt_family="B4A"
                /complement(15667..15707)
                /rpt_family=" (CAA)n"
                /complement(16734..16826)
                /rpt_family=" (TC)n"
                /complement(16826..16855)
                /rpt_family=" (TTCC)n"
                /complement(16863..17036)
                /rpt_family="B2_Mm2"
                /complement(17397..17611)
                /rpt_family="MTD"
                /complement(17612..17757)
```

```
repeat_region /rpt_family="B1_MM"
                /complement(17758..17976)
                /rpt_family="MTD"

Query Match 85.5%; Score 18.8; DB 10; Length 215671;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATCTTGGCCCGCCCTTGAAT 22
   |||||
Db 119097 ACCCTTGGCCCGCCCTTGAAT 119076

RESULT 9
HSA326066/c HSA326066 728 bp DNA linear PRI 18-JUL-2002
LOCUS Homo sapiens genomic sequence surrounding NotI site, clone
DEFINITION NRI-180R.
ACCESSION AJ326066
VERSION AJ326066.1 GI:15870460
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 728)
AUTHORS Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashuba,V.I.,
Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
TITLE NotI flanking sequences: a tool for gene discovery and verification
of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 728)
AUTHORS Zabarovsky,E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden

FEATURES
source
1..728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NRI-180R"

ORIGIN
Query Match 83.6%; Score 18.4; DB 9; Length 728;
Best Local Similarity 95.0%; Pred. No. 6.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAAT 22
   |||||
Db 355 CCGTGGCCCGCCCTTGAAT 336

RESULT 10
AC019617 AC019617 37392 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
DEFINITION AC019617
ACCESSION AC019617
VERSION AC019617.1 GI:6665280
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 37392)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
```

JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10210400 by the submitter. For more information on this record e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1. .37392

Location/Qualifiers

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

ORIGIN

Query Match 83.6%; Score 18.4; DB 2; Length 37392;

Best Local Similarity 95.0%; Pred. No. 5e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20

|||||

Db 2551 ATCCTTGCCCGCCCTTTA 2570

RESULT 11

AC021911/c

LOCUS

DEFINITION Homo sapiens clone RP11-22H4, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC021911

VERSION AC021911.2 GI:9124038

KEYWORDS HTG; HTGS_PHASE0.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 61159)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-22H4

Unpublished

2 (bases 1 to 61159)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Castle,A., Boguslavsky,L., Boukanger,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArillano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6730873.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4122

Center clone name: 22_H_4

* NOTE: This record contains 72 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 743: contig of 743 bp in length

744 843: gap of 100 bp

844 1595: contig of 752 bp in length

1596 1695: gap of 100 bp

1696 2445: contig of 750 bp in length

2446 2545: gap of 100 bp

2546 3303: contig of 758 bp in length

3304 3403: gap of 100 bp

3404 4140: contig of 737 bp in length

4141 4240: gap of 100 bp

4241 4996: contig of 756 bp in length

4997 5096: gap of 100 bp

5097 5845: contig of 743 bp in length

5846 5946: gap of 100 bp

5946 6701: contig of 756 bp in length

6702 7555: contig of 754 bp in length

7556 8396: contig of 741 bp in length

8397 9248: contig of 752 bp in length

9249 9348: gap of 100 bp

9349 10100: contig of 752 bp in length

10101 10200: gap of 100 bp

10201 10951: contig of 751 bp in length

10952 11799: contig of 748 bp in length

11800 11899: gap of 100 bp

11900 12642: contig of 743 bp in length

12643 12742: gap of 100 bp

12743 13476: contig of 734 bp in length

13477 13576: gap of 100 bp

13577 14332: contig of 756 bp in length

14333 14432: gap of 100 bp

14433 15191: contig of 759 bp in length

15192 15291: gap of 100 bp

15292 16036: contig of 745 bp in length

16037 16136: gap of 100 bp

16137 16862: contig of 726 bp in length

16863 17717: contig of 755 bp in length

17718 17817: gap of 100 bp

17818 18532: contig of 715 bp in length

18533 18633: gap of 100 bp

18633 19387: contig of 755 bp in length

19388 19487: gap of 100 bp

19488 20252: contig of 765 bp in length

20253 20352: gap of 100 bp

20353 21099: contig of 747 bp in length

21100 21199: gap of 100 bp

21200 21936: contig of 737 bp in length

21937 22036: gap of 100 bp

22037 22784: contig of 748 bp in length

22785 22884: gap of 100 bp

22885 23638: contig of 754 bp in length

23639 23738: gap of 100 bp

23739 24497: contig of 759 bp in length

24498 24597: gap of 100 bp

24598 25324: contig of 727 bp in length

25325 25424: gap of 100 bp

25425 26176: contig of 752 bp in length

26177 26276: gap of 100 bp

```
* 26277 27031: contig of 755 bp in length
* 27032 27131: gap of 100 bp
* 27132 27893: contig of 762 bp in length
* 27894 27993: gap of 100 bp
* 27994 28762: contig of 769 bp in length
* 28763 28862: gap of 100 bp
* 28863 29620: contig of 758 bp in length
* 29621 29720: gap of 100 bp
* 29721 30462: contig of 742 bp in length
* 30463 30562: gap of 100 bp
* 30563 31316: contig of 754 bp in length
* 31317 31416: gap of 100 bp
* 31417 32152: contig of 736 bp in length
* 32153 32252: gap of 100 bp
* 32253 33011: contig of 759 bp in length
* 33012 33111: gap of 100 bp
* 33112 33849: contig of 738 bp in length
* 33850 33949: gap of 100 bp
* 33950 34717: contig of 768 bp in length
* 34718 34817: gap of 100 bp
* 34818 35559: contig of 742 bp in length
* 35560 35659: gap of 100 bp
* 35660 36412: contig of 753 bp in length
* 36413 36512: gap of 100 bp
* 36513 37253: contig of 741 bp in length
* 37254 37353: gap of 100 bp
* 37354 38106: contig of 753 bp in length
* 38107 38206: gap of 100 bp
* 38207 38958: contig of 752 bp in length
* 38959 39058: gap of 100 bp
* 39059 39810: contig of 752 bp in length
* 39811 39910: gap of 100 bp
* 39912 40664: contig of 754 bp in length
* 40665 40764: gap of 100 bp
* 40765 41493: contig of 729 bp in length
* 41494 41593: gap of 100 bp
* 41594 42335: contig of 742 bp in length
* 42336 42435: gap of 100 bp
* 42436 43181: contig of 746 bp in length
* 43182 43281: gap of 100 bp
* 43282 44015: contig of 734 bp in length
* 44016 44115: gap of 100 bp
* 44116 44871: contig of 756 bp in length
* 44872 44971: gap of 100 bp
* 44972 45729: contig of 758 bp in length
* 45730 45829: gap of 100 bp
* 45830 46590: contig of 761 bp in length
* 46591 47441: contig of 751 bp in length
* 47442 47541: gap of 100 bp
* 47542 48300: contig of 759 bp in length
* 48301 48400: gap of 100 bp
* 48401 49159: contig of 759 bp in length
* 49160 49259: gap of 100 bp
* 49260 50121: contig of 762 bp in length
* 50022 50121: gap of 100 bp
* 50122 50873: contig of 752 bp in length
* 50874 50973: gap of 100 bp
* 50974 51707: contig of 734 bp in length
* 51708 51807: gap of 100 bp
* 51808 52545: contig of 738 bp in length
* 52546 52645: gap of 100 bp
* 52646 53410: contig of 765 bp in length
* 53411 53510: gap of 100 bp
* 53511 54246: contig of 736 bp in length
* 54247 54346: gap of 100 bp
* 54347 55104: contig of 758 bp in length
* 55105 55204: gap of 100 bp
* 55205 55979: contig of 775 bp in length
* 55980 56079: gap of 100 bp
* 56080 56842: contig of 763 bp in length
* 56843 56942: gap of 100 bp
* 56943 57720: contig of 778 bp in length
```

```
* .57721 57820: gap of 100 bp
* 57821 58582: contig of 762 bp in length
* 58583 58682: gap of 100 bp
* 58683 59426: contig of 744 bp in length
* 59427 59526: gap of 100 bp
* 59527 60303: contig of 777 bp in length
* 60304 60403: gap of 100 bp

Query Match      83.6%; Score 18.4; DB 2; Length 61159;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGCCCCGCCCTTGAAT 22
   |||||
Db 55671 CCGTGCCCCGCCCTTGAAT 55652

RESULT 12
AF177767
LOCUS          AF177767          91748 bp DNA linear ROD 24-OCT-2001
DEFINITION    Mus musculus toll-like receptor 4 (Tlr4) gene, Tlr4A allele,
complete cds.
ACCESSION     AF177767
VERSION       AF177767.1 GI:6175874
KEYWORDS      .
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 91748)
AUTHORS       Smirnova,I., Poltorak,A., Chan,E.K.L., McBride,C. and Beutler,B.
TITLE         Phylogenetic variation and polymorphism at the toll-like receptor 4
locus (TLR4)
JOURNAL       Genome Biol. 1 (1), res. 002.1-002.10 (2000)
MEDLINE       20558910
PUBMED        11104518
REFERENCE     2 (bases 1 to 91748)
AUTHORS       Poltorak,A., Smirnova,I. and Beutler,B.
TITLE         Direct Submission
JOURNAL       Submitted (17-AUG-1999) Department of Internal Medicine, University
of Texas, Southwestern Medical Center and the Howard Hughes Medical
Institute, 5323 Harry Hines Boulevard, Dallas, TX 75235-9050, USA

FEATURES             source
     source           1. 91748
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /map="between markers B and D4MIT178"
                     /clone="152C16"
     gene             1. 41297
                     /gene="Tlr4"
                     /allele="Tlr4A"
                     4942. .5509
                     /allele="Tlr4A"
     repeat_region    /rpt_family="L1"
                     complement(13845. .13993)
     repeat_region    /rpt_type="dispersed"
                     complement(19172. .19229)
                     /rpt_family="Alu"
                     /rpt_type="dispersed"
                     join(25960. .26424,32397. .32563,37732. .41297)
     mRNA             /gene="Tlr4"
                     /allele="Tlr4A"
                     /product="Tlr4"
     exon             25960. .26424
                     /evidence="experimental"
                     /gene="Tlr4"
                     /allele="Tlr4A"
     misc_feature      26041
                     /gene="Tlr4"
                     /allele="Tlr4A"
```


/note="alternate transcription start site"
26128..26334
/gene="Tlr4"
/alleles="Tlr4A"
join(26335..26424,32397..32563,37732..39982)
/gene="Tlr4"
/alleles="Tlr4A"
/note="transmembrane subunit of the lipopolysaccharide
(LPS) receptor"
/codon_start=1
/product="toll-like receptor 4"
/protein_id="AAF05317.1"
/db_xref="GI:6175875"
/translation="MPPWLLARTLIMAFSLTPGSLNPCIIEVVPNITVQMDQKL
SKVPDDISSTKNIDLSNPLKILKSYSPNSEIQLWDLSCRETIETDKAWHGLHH
LSNLTGNIPQISFGPSGLTSLENVAVETKGLASLESPIGQILTKLNVANFN
IHSCKLPAVFNLTNVLHVDLSYNTIQTITVNDLOLRENPOVNLSDMSLNPIDIFQ
DOAQGILKHLTLTGNFNSNIMTKICLNLAGLHVRLILGEPKDERNLEIFRPSIM
BGLDVTIDEPLTITNPSDDIVKPHCLANVSAMSLAGVSIKYLEDPVPHKQWLS
TIRCQKQPTDPLTFLSKLTUTMNGKISFKKVALPVSLSYLDLNRNALSFGCCSYIS
DLGTSNLRHLDLSFNGAIIIMSANFLMGNSPKDNTLSNVFANTLTNLTFLDLSKQLYL
DISYNTKIDFGILGTLSTLTKMAGNSPKDNTLSNVFANTLTNLTFLDLSKQLYL
ISWGVDTLHRLQLLNHNNLLFLDSSHYNOLYSLSLTDGSPNRIETSKGLQHPFK
SLAFNLTNNSVACICEKQKPLQWKEOKPLVNVQMTCAFPVEMNTSLVLDNFTK
EDWNRNVLNKEGVPFRPHCLCHLYRDFIPGVIAIANIIQEGFHKSRRKIVVVSRRFI
CWMYKTIISVSIVSVTVAFPIYHFYHLLIAGCKKYSRGSISYDAFVIYSSQN
QSRWCIFYEIAAQWQLSSRSRGIIFIVLEKVKSLRQOVELYRLLSRNTYLEWEDN
PLGRHIFWRLRLKALLDGNAPQETABEEQETATWT"
32397..32563
/gene="Tlr4"
/alleles="Tlr4A"
complement(36099..36345)
/rpt_family="L1"
/rpt_type=dispersed
37732..41297
/gene="Tlr4"
/alleles="Tlr4A"
39985..41297
/gene="Tlr4"
/alleles="Tlr4A"
53942..54016
/rpt_family="Alu"
/rpt_type=dispersed
complement(65648..65858)
/rpt_family="L1"
/rpt_type=dispersed

ORIGIN

Query Match 83.6%; Score 18.4; DB 10; Length 91748;
Best Local Similarity 95.0%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAAT 22
|||||
Db 72159 CCTTGGCCCGCCCTTGAAT 72178
|||||

RESULT 13
AC008212 162876 bp DNA linear INV 17-FEB-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 97A-97B, BAC clone
DEFINITION BACR19L21, complete sequence.
ACCESSION AC008212
VERSION AC008212.8 GI:12957643
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 162876)
REFERENCE Ceiniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,

Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.V., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Scheeler, F.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 97A-97B
Unpublished
2 (bases 1 to 162876)
Ceiniker, S.E., Agbanyani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Kartka, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Masda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 17, 2001 this sequence version replaced gi:7264777.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1..162876
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="97A-97B"
/clone="BACR19L21 (D776)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"

ORIGIN

Query Match 83.6%; Score 18.4; DB 3; Length 162876;
Best Local Similarity 95.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGGCCCGCCCTTGA 20
|||||
Db 29452 ATCCTTGGCCCGCCCTTGA 29471
|||||

RESULT 14
CNS05TDI/163152 bp DNA linear PRI 15-JAN-2003
LOCUS Human chromosome 14 DNA sequence BAC R-97N10 of library RPCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL356800
VERSION AL356800.4 GI:27764220
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163152)

```

AUTHORS
Hellig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissenbach,J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 163152)
Genoscope.
Direct Submission
Submitted (14-JAN-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Jan 15, 2003 this sequence version replaced gi:13897296.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-16013 (AC=ML16153)
Downstream BAC (overlapping the SP6 end) : C-2307P3 -----
Finishing boundaries
FINISHED SEGMENT STARTS AT BASE 1
FINISHED SEGMENT ENDS AT BASE 16490
----- Summary Statistics
----- Assembly program: Phrap; version 2.0
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 :
30 - 39 :
40 - 49 : 254
50 - 59 : 1084
60 - 69 : 2915
70 - 79 : 12350
80 - 89 : 47739
90 - 99 : 98810
-----
Percentage of bases with a quality value >= 40 : 100 %.
-----
FEATURES
source
1. .163152
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-97N10"
/clone_lib="RPCI-11"
146097.146098
/note="CTGATGAATCCCTTAATGATTTTGGTAAAAATCATTAAGTTAAGTGGATA
C
CATCTTGTCATATGATCAAAATGGTTTCGGAAAAATCAATATCAGACAAACAGATGT
G
GAATCGATATTTTACAGCACTCTCTTTTAAACAAATTCGTGCCCGGAATTACACTTAAAC
G
CTCAACAGCTTTAAACGTTGGCTTGGCAGCGCATTAATCTGACTGTAAACCTCTCACTCTTA
C
GAACTTGGCCGTAACTGCCAACCAAGCGAGAACAAACATACATCAACAGCAATCG
A
CGAATTGTTAGGTAATCGTCACTCTTCAAGAGCGACTCGCTGTATACCGTTGGCATG
C
AGCTTTATCTGTTGGGCAATACGATGCCATTTGATCTGTTGACTGCTGTGATATTC
G
GAGCAAAAACGACTTATGTTATTCGAGCTTCAGTCGCACTACACGGTCTGTTCTGTTA
C
CTTTATGAGAAAGCGTTCCTCCGCTTTTCAGAGCAATGTTTCAAGAAAGCTCATGACCAAT
T
CTAGCCGACCTTGGAGCATTTCTACCGAGTAACACACACCGCTCATTTGTCTGATGATG

```

```

C
GGCTTTAAAGTGCCATGGTATATAATCCGTTGAGAAGCTGGGTGGTACTGTTAAGTC
G
GTAAGAGAAAAGTACAAATATGCAGACTAGGAGCGGAAACCTGGAAACCTATCAGCA
A
TTACATGATATGTCATCTAGTCACTCAAGACTTTTAGGCTATAGAGGCTGACTAAAA
G
AATCCAATCTCATGCCAAATTTCTATTGTATATAATCTCGCTCTAAAGGCCGAAAAAATC
A
CGCTCGACACGGACTATTGTGACCACCCGTCACCTAAATCTACTCAGCGTCGGCAA
A
GAGCATGGGTTTCTAGCAACTTAATCTACCTGTTGAAATTCGAAACCAACCAACTTG
T
AATATCTATTGAAAGCGAATGCAGATTGAAGAAACCTTCCGAGACTTGAAGAAGTCTCG
C
TACGGACTAGGCCTACGCCATAGCGGAACGAGCAGCTCAGAGCGTTTTGTATCATGTC
T
CTAATCGCCCTGATGCTTCAACTAACATGTTGGGCTTGGGGCGGTTTCATGCTCAGAAAC
A
GGTTGGGACAAAGCACTTCCAGGCTAACACAGCTCAGAAATCGAAACGTAATCTCTCAACAG
T
CGTTAGGCATGGAAAGTTTTCGGGCATCTCGCTACACAAATAACAAAGGAGACTTAC
T
TGCGTGCAACCCCTACTAGCTCAAAATTTATTCACATGTTTACGCTTTCGCTTTGGGGAAT
T
TGAGGGGATCTCTCAG
/evidence=not_experimental
/insertion_seq="IS10"

ORIGIN
Query Match 83.6%; Score 18.4; DB 9; Length 163152;
Best Local Similarity 95.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAAT 22
Db 60199 CCTTGGCCCGCCCTTGAAT 60180

RESULT 15
AC015945
LOCUS Homo sapiens chromosome 18 clone RP11-106J7 map 18, WORKING DRAFT
DEFINITION AC015945
ACCESSION AC015945
VERSION AC015945.3 GI:7230307
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167642)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhalter,B.,
JOURNAL Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
REFERENCE Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
AUTHORS Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
TITLE Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
JOURNAL Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
REFERENCE Lebecky,J., Liu,C., Locke,K., Macdonald,P., Marcuis,N.,
AUTHORS McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
TITLE Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
JOURNAL Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
REFERENCE Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
AUTHORS Teefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
TITLE Wyman,D., Ye W.J., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6532072.

```


McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Piani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 169892)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campolano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Grant, G., Hagos, B., Headford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Piani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 21, 2000 this sequence version replaced gi:7528391.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9338

Center clone name: 120.D.11

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154793 bases at least Q40

Consensus quality: 162992 bases at least Q30

Consensus quality: 165787 bases at least Q20

Insert size: 153000; agarose-fp

Insert size: 167492; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1034: contig of 1034 bp in length

* 1035 1134: gap of 100 bp

* 1135 2429: contig of 1295 bp in length

* 2430 2529: gap of 100 bp

* 2530 4610: contig of 2081 bp in length

* 4611 4710: gap of 100 bp

* 4711 6383: contig of 1673 bp in length

* 6384 9115: contig of 100 bp

* 9116 9215: contig of 2632 bp in length

* 9216 11823: contig of 100 bp

* 11824 11923: gap of 100 bp

* 11924 13710: contig of 1787 bp in length

* 13711 13810: gap of 100 bp

* 13811 16020: contig of 2210 bp in length

* 16021 16120: gap of 100 bp

* 16121 18663: contig of 2743 bp in length

* 18664 21637: contig of 100 bp

* 21638 21737: gap of 100 bp

* 21738 26182: contig of 4445 bp in length

* 26183 26282: gap of 100 bp

* 26283 30101: contig of 3819 bp in length

* 30102 30201: gap of 100 bp

* 30202 34981: contig of 4780 bp in length

* 34982 35081: gap of 100 bp

* 35082 38927: contig of 3846 bp in length

* 38928 39027: gap of 100 bp

* 39028 44362: contig of 5335 bp in length

* 44363 44462: gap of 100 bp

* 44463 55550: contig of 6088 bp in length

* 55551 56550: gap of 100 bp

* 56551 57140: contig of 6490 bp in length

* 57141 57240: gap of 100 bp

* 57241 64392: contig of 7152 bp in length

* 64393 64492: gap of 100 bp

* 64493 72617: contig of 8125 bp in length

* 72618 72717: gap of 100 bp

* 72718 83466: contig of 10749 bp in length

* 83467 83566: gap of 100 bp

* 83567 96425: contig of 12859 bp in length

* 96426 96525: gap of 100 bp

* 96526 110345: contig of 13819 bp in length

* 110345 110444: gap of 100 bp

* 110445 126599: contig of 16155 bp in length

* 126600 126600: gap of 100 bp

* 126700 14361: contig of 17662 bp in length

* 14362 14461: gap of 100 bp

* 14462 169892: contig of 25431 bp in length.

FEATURES
 source
 1. 169892
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-120D11"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1034
 /note="assembly_fragment"
 1135. 2429
 /note="assembly_fragment"
 2530. 4610
 /note="assembly_fragment"
 4711. 6383
 /note="assembly_fragment"
 6484. 9115
 /note="assembly_fragment"
 9216. 11823
 /note="assembly_fragment"
 11924. 13710
 /note="assembly_fragment"
 13811. 16020
 /note="assembly_fragment"
 16121. 18963
 /note="assembly_fragment"

```

misc_feature      18964..21637
                  /note="assembly_fragment
                  clone_end:17
                  vector_side:right"
misc_feature      21738..26182
                  /note="assembly_fragment"
misc_feature      26283..30101
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:right"
misc_feature      30202..34981
                  /note="assembly_fragment"
misc_feature      35082..38927
                  /note="assembly_fragment"
misc_feature      39028..44362
                  /note="assembly_fragment"
misc_feature      44463..50550
                  /note="assembly_fragment"
misc_feature      50651..57140
                  /note="assembly_fragment"
misc_feature      57241..64392
                  /note="assembly_fragment"
misc_feature      64493..72617
                  /note="assembly_fragment"
misc_feature      72718..83466
                  /note="assembly_fragment"
misc_feature      83567..96425
                  /note="assembly_fragment"

Query Match      83.6%; Score 18.4; DB 2; Length 169892;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAAT 22
    |||
Db 25212 CCGTGGCCCGCCCTTGAAT 25231

RESULT 17
AC008211/c
LOCUS              181904 bp      DNA      linear      INV 31-JUL-2004
DEFINITION        Drosophila melanogaster clone BACR08N16, complete sequence.
ACCESSION         AC008211
VERSION           AC008211.10 GI:50872354
KEYWORDS          HTG.
SOURCE            Drosophila melanogaster (fruit fly)
ORGANISM          Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE          1 (bases 1 to 181904)
AUTHORS            Celniker,S.E., Agbayani,A., Atcaina,T.T., Baxter,E., Blazej,R.G.,
                  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
                  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
                  Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
                  Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
                  Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
                  Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
                  Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
                  Rubin,G.M.
TITLE             Direct Submission
JOURNAL           Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
                  Laboratory, MS 64-121, Berkeley, CA 94720, USA
REFERENCE          2 (bases 1 to 181904)
AUTHORS            Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R.,
                  Hoskins,R., Scapleton,M., Pacleb,J., Park,S., Svirskas,R.,
                  Smith,E., Yu,C. and Rubin,G.
TITLE             Direct Submission
JOURNAL           Submitted (31-JUL-2004) Berkeley Drosophila Genome Project, MS
                  64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
                  Berkeley, CA 94720, US
COMMENT            On Jul 31, 2004 this sequence version replaced gi:13122719.
                  Sequence submitted by:

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This sequence submission incorporates changes made during
reevaluation of the assembly or fingerprint verification of the
clone. For further information about this sequence, including its
location and relationship to other sequences, please visit our
sequence archive web site (http://www.fruitfly.org/sequence/) or
send email to bdgp@fruitfly.org.
Location/Qualifiers
1..181904
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="97A-97A"
/clone="BAC clone BACR08N16 (D775)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACe3.6)"

FEATURES
source
1..181904
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="97A-97A"
/clone="BAC clone BACR08N16 (D775)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACe3.6)"

ORIGIN
Query Match      83.6%; Score 18.4; DB 3; Length 181904;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGGCCCGCCCTTGA 20
    |||
Db 18795 ATCCTTGGCCCGCCCTTGA 18776

RESULT 18
AC018444/c
LOCUS              203381 bp      DNA      linear      HTG 16-MAR-2000
DEFINITION        Homo sapiens chromosome 4 clone RP11-685L5 map 4, WORKING DRAFT
ACCESSION         AC018444
VERSION           AC018444.4 GI:7249141
KEYWORDS          HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 203381)
AUTHORS            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE             Homo sapiens chromosome 4, clone RP11-685L5
JOURNAL           Unpublished
AUTHORS            2 (bases 1 to 203381)
                  Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
                  Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,P.,
                  Boguslavsky,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M.,
                  Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,
                  Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,
                  Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,
                  Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,
                  Karatas,A., Klein,J., Landers,T., Lehoczy,J., Lieu,C., Locke,K.,
                  Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
                  Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,
                  O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,
                  Rochman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,
                  Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
                  Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
                  Zimmer,A. and Zody,M.
TITLE             Direct Submission
JOURNAL           Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome
                  Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT            On Mar 16, 2000 this sequence version replaced gi:6984435.
                  All repeats were identified using RepeatMasker:
                  http://ftp.genome.washington.edu/RM/RepeatMasker.html
                  Smit, A.F.A. & Green, P. (1996-1997)
                  ----- Genome Center
                  Center: Whitehead Institute/ MIT Center for Genome Research

```

```

Center code: WtBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5122
Center clone name: 685 L.5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 192173 bases at least Q40
Consensus quality: 197304 bases at least Q30
Consensus quality: 199244 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 201681; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 797: contig of 797 bp in length
798 897: gap of 100 bp
898 2929: contig of 2032 bp in length
2930 3029: gap of 100 bp
3030 5146: contig of 2117 bp in length
5147 5246: gap of 100 bp
5247 7728: contig of 2482 bp in length
7729 7828: gap of 100 bp
7829 11474: contig of 3646 bp in length
11475 11574: gap of 100 bp
11575 13946: contig of 2372 bp in length
13947 14046: gap of 100 bp
14047 18089: contig of 4043 bp in length
18090 18189: gap of 100 bp
18190 24914: contig of 6725 bp in length
24915 25014: gap of 100 bp
25015 30744: contig of 5730 bp in length
30745 30844: gap of 100 bp
30845 40933: contig of 10089 bp in length
40934 41033: gap of 100 bp
41034 52642: contig of 11609 bp in length
52643 52742: gap of 100 bp
52743 68507: contig of 15765 bp in length
68508 68607: gap of 100 bp
68609 84706: contig of 16099 bp in length
84707 84806: gap of 100 bp
84807 103521: contig of 18715 bp in length
103522 103621: gap of 100 bp
103622 120932: contig of 17311 bp in length
120933 121032: gap of 100 bp
121033 140199: contig of 19167 bp in length
140200 140299: gap of 100 bp
140300 171564: contig of 31265 bp in length
171565 171664: gap of 100 bp
171665 203381: contig of 31717 bp in length.
-----
1. 203381 Location/Qualifiers
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="4"
    /map="4"
    /clone="RP11-685L5"
    /clone_lib="RPC1-11 Human Male BAC"
1. .797
    /note="assembly_fragment
clone end:T7
signature

```


	/rpt_family="TAAAn"	25805..25841
repeat_region	complement(10988..11270)	/rpt_family="LIMEC"
repeat_region	/rpt_family="AluY"	25842..26130
repeat_region	11324..11353	/rpt_family="AluJb"
repeat_region	/rpt_family="AT_rich"	26131..26144
repeat_region	complement(11499..11799)	/rpt_family="LIMEC"
repeat_region	/rpt_family="AluJo"	complement(26145..26442)
repeat_region	complement(12066..12196)	/rpt_family="AluJb"
repeat_region	/rpt_family="L2"	26443..26631
repeat_region	12530..12580	/rpt_family="LIMEC"
repeat_region	/rpt_family="AT_rich"	26632..26947
repeat_region	12607..12822	/rpt_family="AluSx"
repeat_region	/rpt_family="Aluub"	26949..27358
repeat_region	complement(12838..13199)	/rpt_family="LM4"
repeat_region	/rpt_family="LIME3A"	complement(27359..27889)
repeat_region	13200..13484	/rpt_family="MLT2CB"
repeat_region	/rpt_family="AluSq"	27890..27965
repeat_region	13485..13514	/rpt_family="LM4"
repeat_region	/rpt_family="(TAAAn)"	28071..28193
repeat_region	complement(13521..13595)	/rpt_family="LIME3A"
repeat_region	/rpt_family="LIME3A"	complement(28195..28241)
repeat_region	complement(13608..13941)	/rpt_family="(TAAAn)"
repeat_region	/rpt_family="MLT1I"	complement(28242..28526)
repeat_region	complement(14002..14230)	/rpt_family="AluY"
repeat_region	/rpt_family="LIME3A"	
repeat_region	complement(14231..14534)	
repeat_region	/rpt_family="AluY"	
repeat_region	complement(14535..14666)	
repeat_region	/rpt_family="LIME3A"	
repeat_region	complement(14757..15080)	
repeat_region	/rpt_family="LIME3A"	
repeat_region	complement(15165..15362)	
repeat_region	/rpt_family="MIR"	
repeat_region	15499..15735	
repeat_region	/rpt_family="MER46A"	
repeat_region	complement(16289..16341)	
repeat_region	/rpt_family="(CA)n"	
repeat_region	complement(16355..16646)	
repeat_region	/rpt_family="AluSc"	
repeat_region	16690..16764	
repeat_region	/rpt_family="AT_rich"	
repeat_region	16863..17091	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(17724..17827)	
repeat_region	/rpt_family="L2"	
repeat_region	17859..18164	
repeat_region	/rpt_family="AluY"	
repeat_region	complement(18515..18566)	
repeat_region	/rpt_family="(GAAAn)"	
repeat_region	complement(18567..18846)	
repeat_region	/rpt_family="AluSq"	
repeat_region	19382..19672	
repeat_region	/rpt_family="AluJo"	
repeat_region	19674..19975	
repeat_region	/rpt_family="AluY"	
repeat_region	20124..20150	
repeat_region	/rpt_family="AT_rich"	
repeat_region	complement(20157..20465)	
repeat_region	/rpt_family="Aluub"	
repeat_region	20895..21189	
repeat_region	/rpt_family="AluSq"	
repeat_region	21201..21504	
repeat_region	/rpt_family="AluSq"	
repeat_region	complement(21692..22058)	
repeat_region	/rpt_family="L2"	
repeat_region	22478..22783	
repeat_region	/rpt_family="AluY"	
repeat_region	23594..23666	
repeat_region	/rpt_family="AT_rich"	
repeat_region	complement(24497..24800)	
repeat_region	/rpt_family="AluSx"	
repeat_region	25318..25623	
repeat_region	/rpt_family="AluSq"	
repeat_region	/rpt_family="(TAAAn)"	

Query Match	83.6%	Score 18.4;	DB 9;	Length 217421;
Best Local Similarity	95.0%;	Pred. No. 4.5e+02;		
Matches	19;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;

QY	3	CCGTGCCCGCCCTTGGAAT	22
Db	209182	CCGTGCCCGCCCTTGGAAT	209163

RESULT 21	
AP001025	Homo sapiens genomic DNA, chromosome 18 clone:RP11-838N2, complete sequence.
LOCUS	AP001025 218371 bp DNA linear PRI 26-MAR-2003
DEFINITION	Homo sapiens genomic DNA, chromosome 18 clone:RP11-838N2, complete sequence.
ACCESSION	AP001025 GI:29243388
VERSION	AP001025
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE	Homo sapiens genomic DNA
JOURNAL	Published Only in Database (2000)
REFERENCE	2 (bases 1 to 218371)
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan E-mail:hattori@gsc.riken.go.jp, URL:http://hgsp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) On Mar 25, 2003 this sequence version replaced gi:21728138. FEATURES
Location/Qualifiers	1..218371
source	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="18" /map="18p" /clone="RP11-838N2"

Query Match	83.6%	Score 18.4;	DB 9;	Length 218371;
Best Local Similarity	95.0%;	Pred. No. 4.5e+02;		


```
ELPYGLKSLVKEIALVNERIDGVADLRDESDRKGIRVVIDLKRGITIPDIVNAL
YKTPLESIYGLNMLAVNRRPOLLQKLSALSHFLSHRREIIIRIIRYDRYDKAERAH
ILEGLALDNDIDVALLRASATPOEAKDRLMRFGLSEAQSLDRLQLRNLNLE
HKLDEYRELAKKIBYFRSILENIEVLQVMEEFTEIRDTYATKRTVLMDELGG
IAEDILPDDVVITLSRGGYIKRTLDNYQOORGGKIGIAGLHSDDDFVQBFUHTT
NQYLLFTNHHGMHOLKVHVAESRTAKGVHIANLLPLEKDEWTTVTVLSVREPAED
REFLFTKGMVKSASLYARAKGGLIANGLRDDDELIMVREIHDDHDDVLTGCGG
MAIRFACPDVRPMGRSATGVKGVALSGDDRVVACAVLPQDDTSTELMSISSNGYGRK
RIELVLRGSGKGVNFRVTKTGDVIGAMPVDDDEALVLLTSTNKKIRMSVDEVR
VGRATGMVLRDEGAIYVGFDRVDEREVEESAPQSSQEQTAPEGGDAGTEGTAG
"
gene 7835. .8620
CDS /locus_tag="DVU0005"
7835. .8620
/locus_tag="DVU0005"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AAS94489.1"
/db_xref="GI:46447833"
/translation="MRLAILILICILVAACDRQAATEELTEARSAPANREWNDAERLL
EGLRRNPEGAQRAEWANLLVEVHRVNRDERATVELLELTYLNGSNPERAYDILTR
LCELLEAGDRDPRAAEVPRVLELPGLTPTDTRALVSRRLARILVMQRRPPEAEERLKA
CLDTGLSGEGAGCLYDLADLVASEHLDAVDLATTVLGLEGTLPELRSQAGFLA
DALEQGRFAELERFEALRGVYPNDMVIDTRVRLKRLGRK"
complement (8913. .9377)
/locus_tag="DVU0006"
complement (8913. .9377)
/locus_tag="DVU0006"
/notes="identified by match to protein family HMM PF00582"
/codon_start=1
/transl_table=11
/product="universal stress protein family"
/protein_id="AAS94490.1"
/db_xref="GI:46447834"
/translation="MKTEIILPVDGSSSYEAACAAQALCGAEVLLHCVDOTE
QNGEAPLVDTLAEITNNRHKQADALMGATANDLIGPVRDILSDHGIRYMERIVEG
EPATCIPRVAQTRGLIVMGRSGKTPSGRLPVGTVGRVGAACHPVIVAP"
complement (9552. .10937)
/locus_tag="DVU0007"
complement (9552. .10937)
/locus_tag="DVU0007"
/notes="identified by similarity to SP:P17242; match to
protein family HMM PF00152; match to protein family HMM
PF01336; match to protein family HMM TIGR00457"
/codon_start=1
/transl_table=11
/product="asparaginyl-tRNA synthetase"
/protein_id="AAS94491.1"
/db_xref="GI:46447835"
/translation="MQRTQIVDALAATAPQPVIRLCGWRTRDRDAKGSFLEINDGSC
LANIQAVDEGIPAYANSTGAADITGELVESPGKQKQVVRVQTLTLLGAADA
ETYPQKRHSDFIRSAHLARTKTYCAAFIRSEAFATHEFYRERGFFVHTPI
LTGSCGEGAGERNFTVLPVEGSATPAGNRYENDFPKGCNLTVSGOLEAEFTALGL
GKVYTFGTFERAENSTPRHAEFWMIPEVAFADLEDMNLAEQMTTVVRRILDRCL
AADLDFRFDVTLVERLQIADFPFARCSYTEAIELLLKSGKFFPVPVSGDLQIT
EHRYLAEHFQKPVINYPKBIKAFYMRDMDGRTVAAMDVLPVIRGELIGGSGRE
ERLDVLEARIENMGQNLDEYMYWLDLRRFRGSPVHAGFGVGFGLRLMLLTGITNIRDI
PPRFTFGNLEF"
complement (11095. .11220)
gene 83.6%; Score 18.4; DB 1; Length 300129;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 3 CCTTGCCCGCCCTTGAT 22
|||||
DB 167750 CCTTGCCCGCCCTTGAT 167731
```

RESULT 23

BC066993 3172 bp mRNA linear ROD 30-JUN-2004
Mus musculus ubiquitin specific protease 5 (isopeptidase T), mRNA
(cdna clone MGC:91316 IMAGE:5689229), complete cds.

BC066993
MGC: 1 GI:45219890

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3172)

Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 3172)

Strausberg,R.

Direct Submission

Submitted (01-MAR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.

Thomas L. Casavant.

Web site: <http://genome.uiowa.edu>

Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu

Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,

Fiehler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,

Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,

Casavant,T., Soares,M.B.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: Place: Row: Column: 0

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 7305618.

Location/Qualifiers

1. 3172

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="MGC:91316 IMAGE:5689229"

/tissue type="Brain, day 18.5pc (3 -4 kb fragment)"

/clone_lib="NIH_BMAP_EH0p"

/lab host="DH10B"

/note="vector: pYX-ASC"

FEATURES

Source


```
repeat_region /clone lib="RPC1-11"
repeat_region /rpt_family="MER2_type"
misc_feature 1387..1423
misc_feature /rpt_family="AT_rich"
misc_feature 2053..2279
misc_feature /note="similar to Homo sapiens EST BF834955 (NID:g12185705)"
repeat_region 2693..2725
repeat_region /rpt_family="(TTTTA)n"
repeat_region 2697..3000
repeat_region /rpt_family="Alu"
repeat_region 3256..3310
repeat_region /rpt_family="L2"
repeat_region 3403..3559
repeat_region /rpt_family="MeLR"
misc_feature 3529..3883
repeat_region /note="match to EST AA001267 (NID:g1437361) zh82h08.r1"
misc_feature 3795..4856
misc_feature /rpt_family="L1"
repeat_region 4813..4827
repeat_region /note="match to EST BF357078 (NID:g11316152)"
repeat_region 4890..4972
repeat_region /rpt_family="L1"
repeat_region 4973..5334
repeat_region /rpt_family="MeLR"
repeat_region 5335..5695
repeat_region /rpt_family="L1"
misc_feature 5820..6300
misc_feature /note="match to EST BI059057 (NID:g14466587)"
misc_feature 6384..6771
misc_feature /note="match to EST AW799023 (NID:g7850893)"
repeat_region 6479..6525
repeat_region /rpt_family="L2"
repeat_region 6744..6827
repeat_region /rpt_family="L2"
misc_feature 6840..7225
misc_feature /note="match to EST BI493648 (NID:g15332992)"
misc_feature 6901..7388
misc_feature /note="match to EST AA789249 (NID:g2849369) aj27g08.s1"
misc_feature 6942..7437
misc_feature /note="match to EST AA904217 (NID:g3039340) od88e08.s1"
repeat_region 7172..7606
repeat_region /rpt_family="MeLR"
misc_feature 7254..7274
repeat_region /note="match to EST BI493648 (NID:g15332992)"
repeat_region 7902..8212
repeat_region /rpt_family="Alu"
repeat_region 8184..8219
repeat_region /rpt_family="(TAAAAA)n"
repeat_region 8802..9024
repeat_region /rpt_family="Alu"
repeat_region 9287..9577
repeat_region /rpt_family="Alu"
repeat_region 9958..10010
repeat_region /rpt_family="L1"
repeat_region 10019..10106
repeat_region /rpt_family="L1"
repeat_region 10115..10438
repeat_region /rpt_family="Tc2"
repeat_region 10444..10729
repeat_region /rpt_family="Alu"
repeat_region 10742..10802
repeat_region /rpt_family="Tc2"
misc_feature 10832..10833
misc_feature /note="match to EST AW936244 (NID:g8111650)"
misc_feature 11104..11596
misc_feature /note="match to EST AW936244 (NID:g8111650)"
misc_feature 11318..11490
misc_feature /note="match to EST 228747 (NID:g434417)"
misc_feature 11386..11546
misc_feature /note="match to EST BE148119 (NID:g8610843)"
repeat_region 11522..11814
```

```
repeat_region /rpt_family="Alu"
repeat_region 11887..12017
repeat_region /rpt_family="L1"
repeat_region 12027..12187
repeat_region /rpt_family="MIR"
repeat_region 12194..12384
repeat_region /rpt_family="L1"
repeat_region 12547..12669
repeat_region /rpt_family="ERV_L"
repeat_region 13084..13313
repeat_region /rpt_family="MIR"
misc_feature 13286..13292
repeat_region 13470..13720
repeat_region /rpt_family="L2"
repeat_region 14283..14392
repeat_region /rpt_family="Alu"
misc_feature 14443..14680
misc_feature /note="match to EST BF987007 (NID:g12393329)"
repeat_region 14873..15050
repeat_region /rpt_family="L1"
misc_feature 15007..15255
misc_feature /note="match to EST AW995545 (NID:g8255688)"
repeat_region 15044..15071
repeat_region /rpt_family="AT_rich"
repeat_region 15856..15956
repeat_region /rpt_family="Alu"
repeat_region 15957..16268
repeat_region /rpt_family="Alu"
repeat_region 16241..16289
repeat_region /rpt_family="(A)n"
repeat_region 16610..16838
repeat_region /rpt_family="ERV_L"
```

Query Match 80.9%; Score 17.8; DB 9; Length 138435;
Best Local Similarity 90.5%; Pred. NO. 8.e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGCCCCCCCCCTTCAT 22

Db 51014 TCCTGCCCCCCCCCTTCAT 50994

RESULT 30

AC064798/c AC064798 157185 bp DNA linear HTG 17-MAY-2000
LOCUS Homo sapiens chromosome 1 clone RP11-182C20 map 1, WORKING DRAFT
DEFINITION SEQUENCE, 26 unordered pieces.

ACCESSION AC064798

VERSION AC064798.2 GI:7885063

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 157185)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 1, clone RP11-182C20

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 157185)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,K., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,


```

/misc_feature /note="assembly_fragment"
87119..97617
/note="assembly_fragment"
97718..108184
/note="assembly_fragment"
108285..120544
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
120645..137261
/note="assembly_fragment"
137362..157185
/note="assembly_fragment"

ORIGIN
Query Match 80.9%; Score 17.8; DB 2; Length 157185;
Best Local Similarity 90.5%; Pred. No. 8.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGCAAT 22
Dbb 150012 TCCTTGCCCGCCCTTGCAAT 149992

RESULT 31
AC137901 158419 bp DNA linear ROD 30-APR-2004
LOCUS Mus musculus chromosome 6, clone RP23-81E17, complete sequence.
DEFINITION AC137901
ACCESSION AC137901
VERSION AC137901.7 GI:46879273
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 158419)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome 6, clone RP23-81E17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158419)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,P., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 158419)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 158419)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Mihova,T., Mlenga,V.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Matthews,C.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-APR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 30, 2004 this sequence version replaced gi:45592982.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: 122365
Center clone name: 81_E_17
-----
FEATURES
Source
Location/Qualifiers
1..158419
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/map="6"
/clone="RP23-81E17"
/clone_lib="RPCI-23 Female Mouse BAC"
/complement(1..96)
/rpt_family="B1_MM"
1..6
/note="clone boundary"
clone_end:SP6
site:ECORI"
105..127
/rpt_family="TTTTG)n"
complement(128..169)
/rpt_family="B2_Mm1"
180..214
/rpt_family="TTTTG)n"
complement(217..343)
/rpt_family="B1_MM"
repeat_region
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Topham,K.,
Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 158419)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Mihova,T., Mlenga,V.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Matthews,C.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-APR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 30, 2004 this sequence version replaced gi:45592982.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: 122365
Center clone name: 81_E_17
-----
FEATURES
Source
Location/Qualifiers
1..158419
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/map="6"
/clone="RP23-81E17"
/clone_lib="RPCI-23 Female Mouse BAC"
/complement(1..96)
/rpt_family="B1_MM"
1..6
/note="clone boundary"
clone_end:SP6
site:ECORI"
105..127
/rpt_family="TTTTG)n"
complement(128..169)
/rpt_family="B2_Mm1"
180..214
/rpt_family="TTTTG)n"
complement(217..343)
/rpt_family="B1_MM"
repeat_region
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

```
repeat_region 423..470
/rpt_family="B1_MM"
repeat_region 489..602
/rpt_family="B1_MM"
repeat_region 642..783
/rpt_family="B1B"
repeat_region 787..1025
/rpt_family="L1MDa"
repeat_region 1026..1212
/rpt_family="B2_Mm2"
repeat_region 1213..1340
/rpt_family="L1MDa"
repeat_region 1425..1556
/rpt_family="B3"
repeat_region 1557..1675
/rpt_family="B1_MM"
repeat_region 1676..1821
/rpt_family="B1_MM"
repeat_region 1835..2022
/rpt_family="L1MDa"
repeat_region 2027..2243
/rpt_family="B3"
repeat_region 2244..2538
/rpt_family="ID4"
repeat_region 2543..2568
/rpt_family="AT_rich"
repeat_region 2816..2880
/rpt_family="B4"
repeat_region 2881..3093
/rpt_family="B4"
repeat_region 3094..3115
/rpt_family="B4"
repeat_region 3116..3259
/rpt_family="B1_MM"
repeat_region 3260..3498
/rpt_family="B2_Mm1"
repeat_region 3500..3753
/rpt_family="MER58A"
repeat_region 3754..3914
/rpt_family="B4A"
repeat_region 3915..3954
/rpt_family="B4A"
repeat_region 3955..4160
/rpt_family="B4A"
repeat_region 4161..4398
/rpt_family="B3"
repeat_region 4399..4553
/rpt_family="B2_Mm2"
repeat_region 4554..4695
/rpt_family="B1_MM"
repeat_region 4696..4730
/rpt_family="B1_MM"
repeat_region 4731..5114
/rpt_family="CAAAA)n"
repeat_region 5115..5176
/rpt_family="RSINE1"
repeat_region 5177..5387
/rpt_family="CA)n"
repeat_region 5388..5417
/rpt_family="RMR10A"
repeat_region 5418..5444
/rpt_family="RMR10A"
repeat_region 5445..5765
/rpt_family="B3"
repeat_region 5766..6002
/rpt_family="B2_Mm1"
repeat_region 6003..6242
/rpt_family="B3"
repeat_region 6243..7080
/rpt_family="B3"
repeat_region 7081..7394
/rpt_family="MTD"
repeat_region 7395..7533
/rpt_family="B3"

/rpt_family="PB1D9"
7554..7641
/rpt_family="(GAAA)n"
7642..7668
/rpt_family="B4"
complement(7669..7877)
/rpt_family="B2_Mm1"

Query Match 80.9%; Score 17.8; DB 10; Length 158419;
Best Local Similarity 90.5%; Pred. No. 8.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
DB 134843 TCCTTGGCCCGCCCTTGAAT 134863

RESULT 32
AC142254/c
LOCUS AC142254 165264 bp DNA linear HTG 14-AUG-2004
DEFINITION Mus musculus chromosome 6 clone RP24-545G7, WORKING DRAFT SEQUENCE,
6 unordered pieces.
ACCESSION AC142254
VERSION AC142254.3 GI:51241915
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 165264)
AUTHORS Wilson,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165264)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 165264)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 14, 2004 this sequence version replaced gi:29540493.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M.BB0545G07
----- Summary Statistics -----
Sequencing vector: pM13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163461 bases at least Q40
Consensus quality: 164111 bases at least Q30
Consensus quality: 164383 bases at least Q20
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1101: contig of 1101 bp in length
* 1102 1201: gap of unknown length
```

* 1202 2351: contig of 1150 bp in length
 * 2352 2451: gap of unknown length
 * 2452 3694: contig of 1243 bp in length
 * 3695 3794: gap of unknown length
 * 3795 4968: contig of 1174 bp in length
 * 4969 5068: gap of unknown length
 * 5069 22570: contig of 17502 bp in length
 * 22571 165264: contig of 142594 bp in length.
 * 22671 165264: contig of 142594 bp in length.

FEATURES

source
 1..165264
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="6"
 /clone="RP24-545G7"
 1..1101
 /note="assembly_name:Contig11"
 1202..2351
 /note="assembly_name:Contig18"
 2452..3694
 /note="assembly_name:Contig29"
 3795..4968
 /note="assembly_name:Contig42"
 5069..22570
 /note="assembly_name:Contig53"
 22671..165264
 /note="assembly_name:Contig54"
 clone_end:SP6
 vector_side:right"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 165264;

Best Local Similarity 90.5%; Pred. No. 8.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTCCCCCGCCCTTGAAT 22

Db 46543 TCCTTCCCCCGCCCTTGAAT 46523

RESULT 33

AC141993/c
 Locus AC141993 164143 bp DNA linear HTG 21-MAR-2003
 Definition Rattus norvegicus clone CH230-291H5, WORKING DRAFT SEQUENCE, 51
 unordered pieces.

ACCESSION AC141993

VERSION AC141993.1 GI:29135464

KEYWORDS HTG; HTGS, PHASE1; HTGS, DRAFT.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 164143)
 Muzny D, Marie, Metzker M, Lee, Abramson S, Adams C, Alder J, Allen C, Allen H, Alebrooks S, Amin A, Anguiano D, Anyalebechi V, Ayodeji A, Ayodeji M, Baca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F, Biwalto K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon E, Cardenas V, Carter K, Cavazos I, Ceasar H, Center A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L, Davila M, Davis C, Davy-Carroll L, De Anda C, Dederich D, Delgado O, Denson S, Deramo C, Ding Y, Dinh H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Evans K, Egan A, Escoto M, Eugene C, Evans C, Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Frazer C, Gabis A, Ganta R, Garcia A, Garner I, Garza M, Gebregiorgis E, Geer K, Gill R, Grady M, Guerra W, Guevara W, Gunaratne P, Haaland W, Hamill C, Hamilton C, Hamilton K, Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J,

Hernandez R, Hines S, Hladun S, Hodgson A, Hogue M, Hollins B, Howells S, Hulyk S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Karpathy S, Kelly S, Khan Z, King L, Kovar C, Kowik J, Kowik C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorensuhewa L, Louised H, Lozano R, Lu X, Ma J, Maheshwari M, Mahindratne M, Mahmoud M, Malloy K, Mangum A, Mangum B, Mapua P, Martin K, Martin R, Martinez E, Mahoney S, McLeod M, McNeill T, Meenen E, Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Munidasa M, Murphy M, Nair L, Nankervis C, Neal D, Newton N, Nguyen N, Norris S, Nwaokemele O, Okwuonu G, Olarnpunsagoon A, Pal S, Parks K, Pasternak S, Paul H, Perez A, Perez L, Pfannkuch C, Plopper F, Poindexter A, Popovic D, Primus E, Pu L, Puazo M, Quiroz J, Rachlin E, Reeves K, Regier M, Reigh R, Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs F, Rives C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S, Sanders W, Savary G, Scherer S, Scott G, Shatsman S, Shen H, Shetty J, Shvartsbeyn A, Sisson I, Sitter C, Smajs D, Sneed A, Sodergren E, Song X, Z, Sorelle R, Soja J, Steimle M, Strong R, Sutton A, Svatek A, Tabor P, Taylor C, Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Usmani K, Valas R, Vera V, Villasana D, Waldron L, Walker B, Wang J, Wang Q, Wang S, Warren J, Warren R, Wei X, White F, Williams G, Willson R, Wlarczyk R, Wooden H, Worley K, Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V, Yu F, Zhang J, Zhou J, Zhou X, Zhao S, Dunn D, Niederhauser A, Weiss R, Smith D, Holt R, Smith H, Weinstock G, and Gibbs R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 164143)
 Worley K.C.
 Direct Submission
 Submitted (21-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRCT
 Center clone name: CH230-291H5
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 155661 bases at least Q40
 Consensus quality: 162264 bases at least Q30
 Consensus quality: 167406 bases at least Q20
 Estimated insert size: 165492; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 51 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1226: contig of 1226 bp in length
 * 1227 1326: gap of unknown length
 * 1327 2336: contig of 1010 bp in length
 * 2337 2436: gap of unknown length
 * 2437 3494: contig of 1058 bp in length
 * 3495 3594: gap of unknown length
 * 3595 5101: contig of 1507 bp in length

* 5102 5201: gap of unknown length
 * 5202 6937: contig of 1736 bp in length
 * 6938 7037: gap of unknown length
 * 7038 8899: contig of 1862 bp in length
 * 8900 8999: gap of unknown length
 * 9000 10352: contig of 1353 bp in length
 * 10353 10452: gap of unknown length
 * 10454 12315: contig of 1863 bp in length
 * 12316 12415: gap of unknown length
 * 12417 13758: contig of 1343 bp in length
 * 13759 13858: gap of unknown length
 * 13859 15684: contig of 1826 bp in length
 * 15685 15784: gap of unknown length
 * 15785 17132: contig of 1348 bp in length
 * 17133 17232: gap of unknown length
 * 17233 18691: contig of 1459 bp in length
 * 18692 18791: gap of unknown length
 * 18792 20807: contig of 2016 bp in length
 * 20808 20907: gap of unknown length
 * 20909 22625: contig of 1718 bp in length
 * 22626 22725: gap of unknown length
 * 22727 24092: contig of 1367 bp in length
 * 24093 24192: gap of unknown length
 * 24193 25831: contig of 1639 bp in length
 * 25832 25931: gap of unknown length
 * 25933 28222: contig of 2291 bp in length
 * 28223 28322: gap of unknown length
 * 28323 30040: contig of 1718 bp in length
 * 30041 30140: gap of unknown length
 * 30141 32156: contig of 2016 bp in length
 * 32157 32256: gap of unknown length
 * 32257 35282: contig of 3026 bp in length
 * 35283 35382: gap of unknown length
 * 35383 37394: contig of 2012 bp in length
 * 37395 37494: gap of unknown length
 * 37495 39334: contig of 1840 bp in length
 * 39335 39434: gap of unknown length
 * 39435 42439: contig of 3005 bp in length
 * 42440 42539: gap of unknown length
 * 42540 44510: contig of 1971 bp in length
 * 44511 44610: gap of unknown length
 * 44611 47698: contig of 3088 bp in length
 * 47699 47798: gap of unknown length
 * 47799 50879: contig of 3081 bp in length
 * 50880 50979: gap of unknown length
 * 50980 52897: contig of 1918 bp in length
 * 52898 52997: gap of unknown length
 * 52998 55927: contig of 2830 bp in length
 * 55928 55994: contig of 3967 bp in length
 * 55995 59994: gap of unknown length
 * 59995 64016: contig of 4022 bp in length
 * 64017 64116: gap of unknown length
 * 64117 67442: contig of 3526 bp in length
 * 67443 67742: gap of unknown length
 * 67743 70950: contig of 3208 bp in length
 * 70951 71050: gap of unknown length
 * 71051 74056: contig of 3006 bp in length
 * 74057 74156: gap of unknown length
 * 74157 77842: contig of 3686 bp in length
 * 77843 77942: gap of unknown length
 * 77943 81593: contig of 3651 bp in length
 * 81594 81693: gap of unknown length
 * 81694 84557: contig of 2864 bp in length
 * 84558 84657: gap of unknown length
 * 84658 87498: contig of 2841 bp in length
 * 87499 87598: gap of unknown length
 * 87599 91396: contig of 3798 bp in length
 * 91397 91496: gap of unknown length
 * 91497 95893: contig of 4097 bp in length
 * 95894 95994: gap of unknown length
 * 95994 102597: contig of 6903 bp in length
 * 102597 102696: gap of unknown length

* 102697 108349: contig of 5653 bp in length
 * 108350 108449: gap of unknown length
 * 108450 114007: contig of 5558 bp in length
 * 114008 114107: gap of unknown length
 * 114108 120714: contig of 6607 bp in length
 * 120715 120814: gap of unknown length
 * 120815 127396: contig of 6582 bp in length
 * 127397 127496: gap of unknown length
 * 127497 133552: contig of 6056 bp in length
 * 133553 133652: gap of unknown length
 * 133653 141145: contig of 7493 bp in length
 * 141146 141245: gap of unknown length
 * 141246 148028: contig of 6783 bp in length
 * 148029 148128: gap of unknown length
 * 148129 157604: contig of 9476 bp in length
 * 157605 157704: gap of unknown length
 * 157705 166234: contig of 8530 bp in length
 * 166235 166334: gap of unknown length
 * 166335 175892: contig of 9558 bp in length
 * 175893 175992: gap of unknown length
 * 175993 184143: contig of 8151 bp in length.

FEATURES

source
 1. .184143
 /organism="Rattus norvegicus"

Query Match 80.9%; Score 17.8; DB 2; Length 184143;
 Best Local Similarity 90.5%; Pred. No. 8.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGAA 21
 |||||

Db 128905 ATCCTTGCCCGATCCCTTGAA 128885

RESULT 34

DJ293M10

LOCUS Homo sapiens jun dimerization protein gene, partial cds; cfos gene,
 DEFINITION complete cds; and unknown gene.
 ACCESSION AF111167
 VERSION AF111167.2 GI:4572570
 KEYWORDS
 SOURCE Homo sapiens (human)

REFERENCE
 AUTHORS Rowen,L., Madan,A., Qin,S., Ambe,R., Loretz,C., Lasky,S.,
 Ratcliffe,A., Abbasi,N., Dickhoff,R., Shaeffer,T., James,R. and
 Hood,L.
 TITLE Sequencing of human chromosome 14q24.3
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 202267)
 AUTHORS Rowen,L.
 JOURNAL Direct Submission
 TITLE Submitted (07-DEC-1998) Department of Molecular Biotechnology, Box
 357730 University of Washington, Seattle, Washington 98195, USA
 COMMENT On Apr 7, 1999 this sequence version replaced gi:4063506.
 Sequencing methodology: high redundancy shotgun in plasmids.
 Interspersed Repeats were identified with RepeatMasker (available
 from http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

source
 1. .202267
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="14"
 /map="14q24.3"
 /clone="BAC293M10"
 /clone_lib="P. deJong RPC111"

repeat_region 487..652

repeat_region
 rpt_family="MIR"
 complement(843..1140)

```
repeat_region /rpt family="AluSp"
complement(1515. .1818)
repeat_region /rpt family="AluSx"
complement(1846. .2133)
repeat_region /rpt family="AluSx"
complement(2880. .3196)
repeat_region /rpt family="AluSx"
complement(6034. .6338)
repeat_region /rpt family="AluSx"
complement(6341. .6424)
repeat_region /rpt unit="ggaa"
complement(6425. .6725)
repeat_region /rpt family="AluSx"
7013. .7034
repeat_region /rpt family="MIR"
7352. .7561
repeat_region /rpt family="AluSg"
8365. .8670
repeat_region /rpt family="AluSx"
complement(9471. .9770)
repeat_region /rpt family="AluSx"
9937. .10067
repeat_region /rpt family="FLAM_C"
10148. .10286
repeat_region /rpt family="AluSx"
10287. .10574
repeat_region /rpt family="AluSx"
10576. .10742
repeat_region /rpt family="AluSx"
10751. .10812
repeat_region /rpt family="MIR"
complement(10820. .10925)
repeat_region /rpt family="LINE2"
complement(11001. .11208)
repeat_region /rpt family="LINE2"
11394. .11510
repeat_region /rpt family="AluJo"
11715. .11883
repeat_region /rpt family="AluSg"
13081. .13382
repeat_region /rpt family="AluSx"
13425. .13545
repeat_region /rpt family="LINE2"
complement(14610. .14707)
repeat_region /rpt family="LINE2"
complement(14758. .15055)
repeat_region /rpt family="AluSx"
15252. .15553
repeat_region /rpt family="AluSx"
complement(15723. .15766)
repeat_region /rpt family="U2"
15882. .16013
repeat_region /rpt family="AluSp"
16017. .16318
repeat_region /rpt family="AluSx"
16319. .16470
repeat_region /rpt family="AluSg"
16623. .16914
repeat_region /rpt family="AluJo"
complement(16923. .17437)
repeat_region /rpt family="LINE2"
17917. .18216
repeat_region /rpt family="AluSx"
19484. .19779
repeat_region /rpt family="AluJo"
complement(19793. .19867)
repeat_region /rpt family="LINE2"
complement(21266. .21565)
repeat_region /rpt family="AluSx"
22090. .22181
repeat_region /rpt family="GC_rich"
22960. .23012
repeat_region /rpt unit="taaa"

complement(23388. .23429)
/rpt family="GC_rich"
join(23969. .24109,24863. .25114,25546. .25653,25769. .26410)
/notes="oncogene; Intron-exon boundaries defined in
relation to cDNA in K00650"
/codon_start=1
/product="cfos"
/protein_id="AAC98315.1"
/db_xref="GI:4063509"
/translation="MMPGFNADYEASSRSCSSASPGDSLVSYPHSPADSPSSMGSPV
NAQDCTDLAVSSANFIPVTVAISTPDLQMLVQPALVSSVAPQOTRAPHPFGVPAPS
AGAYSRAVGVKMTGGRAQSIGRRKVEQLSPSEEEKRRIRRRNKMAAKCRNRRE
LTDLQAEFTDQLEDEKALQTEIANLLKEKEKLFILAAHRPAKIPDGLGPEEMSV
ASLDLTGCLPEVATPESEEAFTPLNDPEPKPVBPVKISISSMELEKTEPDDFLFPA
SSRPSGSETARSVPMDLSGSGFYAADMEPLHSGSLGMPGPMATELEPLCTPVVTCPTPC
TAYTSSPVFTYPEADSPSCAAHRKGSSSNEPSSDSLSSTLLAL"
complement(27038. .27113)
/rpt unit="taaaa"
27775. .28065
/rpt family="AluSp"
28922. .29193
/rpt family="AluJb"
complement(29678. .29955)
/rpt family="AluSx"
30039. .30340
/rpt family="AluSg"
30440. .30536
/rpt family="LINE2"
31139. .31167
/rpt family="POLY_A"
31274. .31343
/rpt family="LTR9"
31385. .31716
/rpt family="LINE2"
31738. .32032
/rpt family="AluSx"
32038. .32163
/rpt family="LINE2"
32634. .32673
/rpt unit="cat"
complement(32983. .33266)
/rpt family="AluSx"
complement(33646. .33935)
/rpt family="AluSx"
34003. .34178
/rpt family="MIR"
34179. .34480
/rpt family="AluSg"
complement(34553. .34604)
/rpt family="LINE2"
complement(35102. .35387)
/rpt family="LINE2"
35589. .35632
/rpt family="LINE2"
complement(35635. .35824)
/rpt family="MIR"
complement(36196. .36778)
/rpt family="MIR"
complement(join(<37092. .37350,39157. .>39299))
/product="unknown"
complement(join(37092. .37350,39157. .39299))
/notes="similar to pH 34; This CDS is predicted by Genscan.
There are no ESTs that match this sequence. BLASTX
indicates similarity to mouse retinoic acid regulatory
protein pH 34"
/codon_start=1
/product="unknown"
/protein_id="AAC98314.1"
/db_xref="GI:4063508"
/translation="MGRARAGRHRGAGWGPCGAVTCVADGAYVSPKPNKORTRE
LLAPRPWIIPIYIEQMSKAMQLKALSSDITEVVYSSYIKLQTKMQLQSMAEWHC
HQEQGMLKCAEAMNALKLDPMKRTSFRPM"
complement(37853. .38149)
repeat_region
```

```

repeat_region      /rpt family="AluJb"
38315..38355
/rpt unit="ca"
repeat_region      complement(38356..38532)
/rpt family="MIR"
38618..38895
/rpt family="AluX"
39236..39284
/rpt family="GC rich"
39316..39386
/rpt family="GC rich"
complement(40471..40542)
/rpt family="MIR"
41550..41849
/rpt family="AluX"
complement(41918..42216)
/rpt family="AluX"

Query Match      80.9%; Score 17.8; DB 9; Length 202267;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TCCTGCCCGCCGCTTGAAT 22
Db      39372 TCCTGCCCGCCGCTTGAAT 39392

RESULT 35
AC120203
LOCUS      AC120203      207930 bp      DNA      linear      HTG 23-NOV-2002
DEFINITION      Pan troglodytes clone RP43-134H13, WORKING DRAFT SEQUENCE, 10
ordered pieces.
ACCESSION      AC120203
VERSION      AC120203.5 GI:25188941
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      1 (bases 1 to 207930)
AUTHORS      Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Marquies,E.H., Masello,C., Maskeri,B., McDowell,J.,
Paquirigan,C., Pearson,R., Portnov,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schueler,W.G., Sison,C.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 207930)
Green,E.D.
Direct Submission
Submitted (04-MAY-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 207930)
Green,E.D.
Direct Submission
Submitted (23-NOV-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Nov 23, 2002 this sequence version replaced gi:23308023.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: ddb
Center Clone name: 134H13

```

The sequence data in this record represents an 'enhanced'

version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 205807 bases at least Q40
Consensus quality: 206425 bases at least Q30
Consensus quality: 206844 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 207030; sum-of-contigs
Quality coverage: 9.42x in Q20 bases; agarose-fp
Quality coverage: 8.64x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 301: contig of 301 bp in length
* 302 401: gap of unknown length
* 303 3102: contig of 2601 bp in length
* 3103 3102: gap of unknown length
* 16979 16978: contig of 13876 bp in length
* 17079 6095: contig of 49017 bp in length
* 66096 66195: gap of unknown length
* 83082 83082: contig of 16887 bp in length
* 83083 83182: gap of unknown length
* 83183 100224: contig of 16942 bp in length
* 100125 100224: gap of unknown length
* 100225 135118: contig of 34894 bp in length
* 135119 135218: gap of unknown length
* 135219 162500: contig of 27282 bp in length
* 162501 162600: gap of unknown length
* 162601 189441: contig of 26841 bp in length
* 189442 189541: gap of unknown length
* 189542 207930: contig of 18389 bp in length.

FEATURES

	Location/Qualifiers	source
	1..207930	
	/organism="Pan troglodytes"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9598"	
	/clone="RP43-134H13"	
	/clone_lib="RP43"	
misc_feature	1..301	
	/note="assembly_fragment"	
	clone_end:T7	
	vector_side:left	
misc_feature	402..3002	
	/note="assembly_fragment"	
misc_feature	3103..16978	
	/note="assembly_fragment"	
misc_feature	17079..66095	
	/note="assembly_fragment"	
misc_feature	66196..83082	
	/note="assembly_fragment"	
misc_feature	83183..100124	
	/note="assembly_fragment"	
misc_feature	100225..135118	

```

/note="assembly_fragment"
135219. .162500
/note="assembly_fragment"
162601. .189441
/note="assembly_fragment"
189542. .207930
/note="assembly_fragment
clone end:SP6
vector_side:right"

ORIGIN
Query Match      80.9%; Score 17.8; DB 2; Length 207930;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGGAAT 22
Db 159750 TCCTTGCCCGCCCTTGGAAT 159770

RESULT 36
AC097047
LOCUS
DEFINITION
AC097047 214493 bp DNA linear HTG 10-MAY-2003
Rattus norvegicus clone CH230-175H22, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
AC097047
AC097047.6 GI:30520966
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 214493)
Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A.,
Karpman,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karkathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhea,L., Louisegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelmehe,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puafo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

```

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trzjos, Z., Umani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 214493)
Worley, K.C.
Direct Submission
Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 214493)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:23664796.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GICC
Center clone name: CH230-175H22
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 204101 bases at least Q40
Consensus quality: 206739 bases at least Q30
Consensus quality: 208781 bases at least Q20
Estimated insert size: 215171, sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 44099: contig of 44099 bp in length
* 44100 44199: gap of unknown length
* 44200 80379: contig of 36180 bp in length
* 80380 80479: gap of unknown length
* 80480 206793: contig of 126314 bp in length
* 206794 206893: gap of unknown length
* 206894 207972: contig of 1079 bp in length
* 207973 214493: gap of unknown length
* 208073 214493: contig of 6421 bp in length.
Location/Qualifiers

FEATURES


```

source      1. 214493
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-175H22"
misc_feature 1505..2087
/notes="clone_boundary"
clone_end:Sp6
site:EcoRI
end_sequence:BH354806"
misc_feature 41254..42548
/notes="wgs_contig"
misc_feature 42599..44099
/notes="wgs_contig"
misc_feature 44200..45888
/notes="wgs_contig"
misc_feature 66958..106079
/notes="clone_boundary"
clone_end:T7
site:EcoRI
end_sequence:BH354805"
ORIGIN
Query Match      80.9%; Score 17.8; DB 2; Length 214493;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCTTGGCCCGCCCTTGA 21
Db 189604 ATCTTCCCTCCCTTGA 189624

RESULT 37
AC095975
LOCUS      AC095975      226406 bp      DNA      linear      HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-11L19, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
ACCESSION AC095975
VERSION    AC095975.6 GI:24941076
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 226406)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kwile,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,K., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoileme,O., Okwuonu,G., Olarnpunagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajic,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Soza,J.,
Steimle,M., Strong,N., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Weinhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 226406)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 226406)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:2272500.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly ('a' contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GERA
Center clone name: CH230-11L19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 20687 bases at least Q40
Consensus quality: 20597 bases at least Q30
Consensus quality: 208962 bases at least Q20
Estimated insert size: 210996; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

* as soon as it is available and the accession number will
* be preserved.

* 1 40908: contig of 40908 bp in length
* 40909 41008: gap of unknown length
* 41009 96121: contig of 55113 bp in length
* 96122 96221: gap of unknown length
* 96222 226406: contig of 130185 bp in length.

FEATURES

Location/Qualifiers

1..226406
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-111L19"
1..1279
/note="wgs_contig"
14893..17096
/note="wgs_contig"
158804..159971
/note="wgs_contig"
202126..204418
/note="wgs_contig"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 226406;
Best Local Similarity 90.5%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGTGCCCGCCCTTGAAT 22

Db 17597 TCCTGTGCCCGCCCTTGAAT 17617

RESULT 38

AC114469

LOCUS

DEFINITION Rattus norvegicus clone CH230-5903, WORKING DRAFT SEQUENCE, 10 linear HTG 19-NOV-2002
unordered pieces.

AC114469

VERSION AC114469.5 GI:25072896

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 227369)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biewald,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dimo,K., Divya,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,S., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarpunsaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 227369)

Worley,K.C.

Direct Submission

Submitted (09-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 227369)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23269887.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTJO

Center clone name: CH230-5903

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 213596 bases at least Q40

Consensus quality: 215947 bases at least Q30

Consensus quality: 217473 bases at least Q20

Estimated insert size: 218400; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: this is a 'working draft' sequence. It currently

* consists of 10 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

1 30701: contig of 30701 bp in length
* 30702 30801: gap of unknown length
* 30802 30899: contig of 37098 bp in length
* 67900 67999: gap of unknown length
* 68000 194823: contig of 126824 bp in length
* 194824 194923: gap of unknown length
* 194924 195972: contig of 1049 bp in length
* 195973 196072: gap of unknown length
* 196073 197161: contig of 1089 bp in length
* 197162 197261: gap of unknown length
* 197262 198531: contig of 1270 bp in length
* 198532 198631: gap of unknown length
* 198632 201116: contig of 2485 bp in length
* 201117 202116: gap of unknown length
* 202117 202649: contig of 1433 bp in length
* 202650 202749: gap of unknown length
* 202750 208642: contig of 5893 bp in length
* 208643 208742: gap of unknown length
* 208743 227369: contig of 18627 bp in length.

FEATURES

source
1. .227369
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-5903"
misc_feature
68000..69071
/note="wgs_contig"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 227369;
Best Local Similarity 90.5%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCTCCCTCCCTTCAA 21

Db 177224 ATCTTCCCTCCCTCCCTTCAA 177244

RESULT 39

AC002397

LOCUS AC002397 227538 bp DNA linear ROD 17-SEP-1998
DEFINITION Mouse chromosome 6 BAC-284H12 (Research Genetics mouse BAC library)
complete sequence.

ACCESSION AC002397 AC002393

VERSION AC002397.1 GI:3287367

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM

1 (bases 1 to 227538)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Ansari-Lari, M.A., Oeltjen, J.C., Schwartz, S., Zhang, Z., Muzny, D.M.,
Lu, J., Gorrell, J.H., Chinault, A.C., Belmont, J.W., Miller, W. and
Gibbs, R.A.

Comparative sequence analysis of a gene-rich cluster at human
chromosome 12p13 and its syntenic region in mouse chromosome 6
Genome Res. 8 (1), 29-40 (1998)

98112780

9445485

REMARK

M. Ali Ansari-Lari, John C. Oeltjen, Donna M. Muzny, Jing Lu, James
H. Gorrell, A. Craig Chinault, and John W. Belmont are affiliated
with the Baylor College of Medicine. Scott Schwartz, Zheng Zhang,
and Webb Miller are affiliated with the Department of Computer
Science and Engineering at Penn State University, University Park,
Pennsylvania 16802.

2 (bases 1 to 227538)

Muzny, D., Aronson, A.D., Adams, C., Brundage, E., Bunac, C.,
Carvelli, K., Chacko, J., Chen, J., Di, W., Ding, Y., Dugan, S.,
Durbin, J., Forcum, J., Ganesh, R., Garcia, C., Goodman, M.,
Gorrell, J.H., Hayward, M., Hernandez, J., Jackson, L., Jin, S.,
Kampal, R., Karpathy, S., Kovar, C., Leal, B., Li, Y., Lichtarge, O.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
3 (bases 1 to 93626)

Muzny, D., Ansari-Lari, M.A., Timms, K.M., Yu, W., Dugan, S., Lu, J.,
Shen, Y., Rowland, K., Liu, W., Perez, L., Ding, Y., Hayward, M.,
Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,
Worley, K., Chen, E., Forcum, J., Aronson, A.D., Chiu, M.W.,
Gorrell, J.H., Brundage, E., Di, W., Chinault, C., Nelson, D. and
Gibbs, R.A.

Direct Submission

Unpublished

4 (bases 1 to 227538)

Chiu, M.W.

Direct Submission

Submitted (01-AUG-1997) Molecular and Human Genetics, Baylor

College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 227538)

Worley, K.C.

Direct Submission

Submitted (13-JUN-1998) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 227538)

Worley, K.C.

Direct Submission

Submitted (02-JUL-1998) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

7 (bases 1 to 227538)

Worley, K.C.

Direct Submission

Submitted (08-JUL-1998) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

8 (bases 1 to 227538)

Worley, K.C.

Direct Submission

Submitted (17-SEP-1998) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 2, 1998 this sequence version replaced gi:3219328.

Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by

Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.

FEATURES

source

1. .227538
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/clone="BAC-284H12"

repeat_region
45..169

repeat_region
175..212

repeat_region
complement(258..517)

/rpt_family="PB1D9"

/rpt_family="(CAAA)n"

/complement(258..517)

Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puaso, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajis, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, S., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 234597)
Worley, K.C.

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 234597)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:22856199.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXDT
Center clone name: CH230-197M4
----- Summary Statistics
Assembly program: Phrap, version 0.990129
Consensus quality: 217574 bases at least Q40
Consensus quality: 219194 bases at least Q30
Consensus quality: 219929 bases at least Q20
Estimated insert size: 221965; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 234597: contig of 234597 bp in length.

FEATURES
source
1..234597
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-197M4"
1..1010
/note="wgs contig"
233563..234597
/note="wgs_contig"

misc_feature
misc_feature
ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 234597;
Best Local Similarity 90.5%; Pred No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCCCGCCCTTGAAT 22
|||||
DB 117256 TCCTTGGCCCCCGCCCTTGAAT 117236

RESULT 41
AC120627
LOCUS
DEFINITION
Rattus norvegicus clone CH230-71L23, *** SEQUENCING IN PROGRESS
***, 7 unordered pieces.
AC120627
AC120627.4 Gi:23267616
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 257605)
Murny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alébrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshwama, L., Loulseg, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, P., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

source

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0; Mismatches

2; Indels

0; Gaps

RESULT 42

AC096060/c

LOCUS

DEFINITION

AC096060.6

GI:24817927

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

* 251328 251427: gap of unknown length
 * 251428 251440: contig of 1013 bp in length
 * 252441 252450: gap of unknown length
 * 252541 253608: contig of 1068 bp in length
 * 253609 253708: gap of unknown length
 * 253709 253721: contig of 1613 bp in length
 * 253822 25421: gap of unknown length
 * 255422 257605: contig of 2184 bp in length.

FEATURES

Location/Qualifiers

1..257605

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-71L23"

2552..3306

/note="clone_boundary"

clone_end:T7

site:EcoRI

end_sequence:BH299627"

193098..194564

/note="wgs_contig"

245265..246953

/note="wgs_contig"

Score 17.8; DB 2; Length 257605;

90.5%; Pred. No. 8.4e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22

|||||

Db 206777 TCCTTGCCCGCCCTTGAAT 206757

RESULT 42

AC096060/c

LOCUS

DEFINITION

AC096060.6

GI:24817927

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 263879)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, J., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, C., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwakenah, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 263879)
Worley, K.C.

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:2777237.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GEIK

Center clone name: CH230-42N9

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 245233 bases at least Q40

Consensus quality: 247905 bases at least Q30

Consensus quality: 250127 bases at least Q20

Estimated insert size: 249743; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 244902: contig of 244902 bp in length
* 244903 245002: gap of unknown length
* 245003 249201: contig of 4199 bp in length
* 249202 249301: gap of unknown length
* 249302 250381: contig of 1080 bp in length
* 250382 250481: gap of unknown length
* 250482 251570: contig of 1089 bp in length
* 251571 251670: gap of unknown length
* 251671 252006: contig of 1236 bp in length
* 252007 253006: gap of unknown length
* 253007 254941: contig of 1934 bp in length
* 254941 255041: gap of unknown length
* 255041 256850: gap of unknown length
* 256851 258257: contig of 1407 bp in length
* 258258 258357: gap of unknown length
* 258358 259762: contig of 1305 bp in length
* 259763 261986: gap of unknown length
* 261987 262086: gap of unknown length
* 262087 263879: contig of 1793 bp in length.

FEATURES
source

1..263879
/organism="Rattus norvegicus"
/mol_type="Genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-42N9"

misc_feature

1..1187

misc_feature

12505..14384

misc_feature

244179..244893

/note="wgs_contig"
/note="wgs_contig"
/note="clone_boundary"
clone_end:T7
site:EcoRI
end sequence:BH261657"
245003..246868
/note="wgs_end_extension"
clone_end:T7"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 263879;
Best Local Similarity 90.5%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGAA 21

|||||

Db 234083 ATCCTTGCCAGTCCCTTGAA 234063

RESULT 43

AC120979

LOCUS

DEFINITION Rattus norvegicus clone CH230-39C7, WORKING DRAFT SEQUENCE, 3
unordered pieces.

ACCESSION

AC120979

VERSION

AC120979.4 GI:25137857

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 267829)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, W., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Esgueta-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvate,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Huliyk,S., Hume,J., Idelbird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,D.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,S., Song,X.-Z., Sorelle,R., Soosa,J., Stealmle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
2 (bases 1 to 267829)
Worley,K.C.
Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 267829)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23195545.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMR
Center clone name: CH230-39C7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 220303 bases at least Q40
Consensus quality: 222409 bases at least Q30
Consensus quality: 223823 bases at least Q20
Estimated insert size: 223136; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 10577: contig of 10577 bp in length
* 10578 10677: gap of unknown length
* 10678 265487: contig of 255810 bp in length
* 265488 265587: gap of unknown length
* 265588 267829: contig of 1242 bp in length.
----- Location/Qualifiers
source
1..267829
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-39C7"
1..11172
/note="wgs_contig"
1
/note="clone_boundary"
clone_end:T7
site:
end sequence:BH283186"
45574..46841
/note="wgs_contig"
complement(247964..248822)
/note="clone_boundary"
clone_end:Sp6
site:
end sequence:BH283188"
261087..262110
/note="wgs_end_extension"
clone_end:Sp6"
264450..265487
/note="wgs_end_extension"
clone_end:Sp6"

Query Match 80.9%; Score 17.8; DB 2; Length 267829;
Best Local Similarity 90.5%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCGCTTGAAT 22
|||||
DB 147287 TCCTTGCCCGCCCGCTTGAAT 147307
|||||
RESULT 44
AC125633/c
LOCUS Rattus norvegicus clone CH230-1086, *** SEQUENCING IN PROGRESS ***
DEFINITION AC125633
ACCESSION AC125633.2 GI:23264559
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 267946)
AUTHORS	Muzny, D. Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
REFERENCE	2 (bases 1 to 267946)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 267946)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Sep 21, 2002 this sequence version replaced gi:21629120. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
Center:	Baylor College of Medicine
Center code:	BCM
Web site:	http://www.hgsc.bcm.tmc.edu/
Contact:	hgsc-help@bcm.tmc.edu
Project Information	
Center project name:	GDQW
Center clone name:	CH230-10P6
Summary Statistics	
Assembly program:	Phrap; version 0.990329
Consensus quality:	225690 bases at least Q40
Consensus quality:	227722 bases at least Q30
Consensus quality:	228860 bases at least Q20
Estimated insert size:	256319; sum-of-contigs estimation
Quality coverage:	3x in Q20 bases; sum-of-contigs estimation
NOTE:	Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
NOTE:	This sequence may represent more than one clone.
NOTE:	This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
This sequence will be replaced	
by the finished sequence as soon as it is available and the accession number will be preserved.	
1 267946: contig of 267946 bp in length.	
Location/Qualifiers	
1..267946	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-10P6"
1..1874	/note="wgs end extension"
	clone_end:Sp6"
2753..3400	/note="clone boundary"
	clone_end:Sp6"
	site:EcoRI
	end_sequence:BH305497"
	complement(266191..266656)
	/note="clone boundary"
	clone_end:T7
	site:EcoRI
	end_sequence:BH305496"
ORIGIN	
Query Match	80.9%; Score 17.8; DB 2; Length 267946;
Best Local Similarity	90.5%; Pred. No. 8.4e+02;
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	2 TCCTTGCCCGCCCTTGAAAT 22
Db	253479 TCCTTGCCCGCCCTTGAAAT 253459
RESULT 45	
AC113707	
LOCUS	AC113707 274467 bp DNA linear HTG 20-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-102K5, *** SEQUENCING IN PROGRESS
	***, 3 unordered pieces.
ACCESSION	AC113707
VERSION	AC113707.5 GI:25138828
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE AUTHORS

1 (bases 1 to 274467)
Munzy,D.,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.D., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li-Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwackelmech,O., Okwono,G., Olarnpungoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Reigh,R., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,I., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanai,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

TITLE JOURNAL

2 (bases 1 to 274467)
Unpublished
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE AUTHORS

3 (bases 1 to 274467)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 20, 2002 this sequence version replaced gi:22856688.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSBDB
Center clone name: CH230-102K5
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 216922 bases at least Q40
Consensus quality: 221334 bases at least Q30
Consensus quality: 224294 bases at least Q20
Estimated insert size: 223028; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 121482: contig of 121482 bp in length
* 121483 121582: gap of unknown length
* 121583 272741: contig of 151159 bp in length
* 272742 272841: gap of unknown length
* 272842 274467: contig of 1626 bp in length.

FEATURES source

Location/Qualifiers
1..274467
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-102K5"
misc_feature 1..1164
/note="wgs_contig"
misc_feature 55460..56868
/note="wgs_contig"
misc_feature 121583..123218
/note="wgs_contig"
misc_feature 151147..153781
/note="wgs_contig"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 274467;
Best Local Similarity 90.5%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCTCCCGCCCTTGAA 21

Db 45056 ATCCTTGCCTCCCGCCCTTGAA 45076

RESULT 46

AC073666/c
LOCUS
DEFINITION Mus musculus clone CT7-236L20, WORKING DRAFT SEQUENCE, 102
ACCESSION AC073666
VERSION AC073666.1 GI:8810283
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
JOURNAL
AUTHORS
TITLE
JOURNAL
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 336873)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 336873)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1044089
Center Clone name: RG-MBAC_236120

Summary Statistics
Consensus quality: 258685 bases at least Q40
Consensus quality: 297146 bases at least Q30
Consensus quality: 306190 bases at least Q20
Estimated insert size: 177000; pulse field gel estimation
Quality coverage: 9.99 in Q20 bases; sum-of-contigs estimation
Quality coverage: 5.41 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 102 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1022: contig of 1022 bp in length
* 1023 1122: gap of unknown length
* 1123 2222: contig of 1100 bp in length
* 2223 2322: gap of unknown length
* 2323 3362: contig of 1040 bp in length
* 3363 3462: gap of unknown length
* 3463 4572: contig of 1110 bp in length
* 4573 4672: gap of unknown length
* 4673 5777: contig of 1105 bp in length
* 5778 5877: gap of unknown length
* 5878 7059: contig of 1182 bp in length
* 7060 7159: gap of unknown length
* 7160 8819: contig of 1660 bp in length
* 8820 8919: gap of unknown length
* 8920 10510: contig of 1591 bp in length
* 10511 11765: contig of 1155 bp in length
* 11766 11865: gap of unknown length
* 11866 12970: contig of 1105 bp in length
* 12971 13070: gap of unknown length
* 13071 14161: contig of 1091 bp in length
* 14162 14261: gap of unknown length
* 14262 15776: contig of 1515 bp in length
* 15777 15876: gap of unknown length
* 15877 17078: contig of 1202 bp in length
* 17079 17178: gap of unknown length
* 17179 18434: contig of 1256 bp in length
* 18435 19654: gap of unknown length
* 19655 19754: gap of unknown length
* 19755 21830: contig of 2076 bp in length
* 21831 21930: gap of unknown length
* 21931 22991: contig of 1061 bp in length
* 22992 23091: gap of unknown length
* 23092 24522: contig of 1431 bp in length
* 24523 24622: gap of unknown length
* 24623 25909: contig of 1287 bp in length

25910 26009: gap of unknown length
26010 27445: contig of 1336 bp in length
27446 27446: gap of unknown length
27446 28729: contig of 1284 bp in length
28730 28829: gap of unknown length
28830 30030: contig of 1201 bp in length
30031 30130: gap of unknown length
30131 31493: contig of 1363 bp in length
31494 31593: gap of unknown length
31594 32743: contig of 1149 bp in length
32743 32842: gap of unknown length
32843 34509: contig of 1667 bp in length
34510 35838: gap of unknown length
35839 35938: contig of 1229 bp in length
35939 37640: contig of 1702 bp in length
37641 37740: gap of unknown length
37741 38914: contig of 1174 bp in length
38915 39014: gap of unknown length
39015 40280: contig of 1266 bp in length
40281 40380: gap of unknown length
40381 41443: contig of 1063 bp in length
41444 41543: gap of unknown length
41544 42740: contig of 1197 bp in length
42741 42840: gap of unknown length
42841 44578: contig of 1738 bp in length
44579 44578: gap of unknown length
44579 45080: contig of 1230 bp in length
45081 46008: gap of unknown length
46009 47243: contig of 1235 bp in length
47244 47343: gap of unknown length
47344 48612: contig of 1269 bp in length
48613 48712: gap of unknown length
48713 50500: contig of 1788 bp in length
50501 50600: gap of unknown length
50601 52164: contig of 1564 bp in length
52165 52264: gap of unknown length
52265 53302: contig of 1038 bp in length
53303 53402: gap of unknown length
53403 54966: contig of 1564 bp in length
54967 55066: gap of unknown length
55067 56399: contig of 1333 bp in length
56400 56499: gap of unknown length
56500 57882: contig of 1383 bp in length
57883 57982: gap of unknown length
57983 59646: contig of 1664 bp in length
59647 59746: gap of unknown length
59747 61226: contig of 1480 bp in length
61227 61326: gap of unknown length
61327 62730: contig of 1404 bp in length
62731 62830: gap of unknown length
62831 64984: contig of 2154 bp in length
64985 65084: gap of unknown length
65085 66497: contig of 1413 bp in length
66498 66597: gap of unknown length
66599 68989: contig of 2392 bp in length
68990 69089: gap of unknown length
69090 71064: contig of 1975 bp in length
71065 71164: gap of unknown length
71165 73049: contig of 1885 bp in length
73050 73149: gap of unknown length
73150 74802: contig of 1653 bp in length
74803 74902: gap of unknown length
74903 77021: contig of 2119 bp in length
77022 77121: gap of unknown length
77122 79133: contig of 2012 bp in length
79134 79233: gap of unknown length
81483 81483: contig of 2250 bp in length
81584 82723: gap of unknown length
82724 82823: contig of 1140 bp in length
82824 84858: gap of unknown length
84859 84958: contig of 2035 bp in length

```
* 84959 86160: contig of 1202 bp in length
* 86161 86260: gap of unknown length
* 86261 88047: contig of 1787 bp in length
* 88048 88147: gap of unknown length
* 88148 89161: contig of 1014 bp in length
* 89162 89261: gap of unknown length
* 89262 90639: contig of 1378 bp in length
* 90640 90739: gap of unknown length
* 90740 91932: contig of 1193 bp in length
* 91933 92032: gap of unknown length
* 92033 94217: contig of 2185 bp in length
* 94218 94317: gap of unknown length
* 94318 96007: contig of 1690 bp in length
* 96008 96107: gap of unknown length
* 96108 97540: contig of 1433 bp in length
* 97541 97640: gap of unknown length
* 97641 100523: contig of 2883 bp in length
* 100524 100623: gap of unknown length
* 100624 103713: contig of 3090 bp in length
* 103714 103813: gap of unknown length
* 103814 106447: contig of 2634 bp in length
* 106448 106547: gap of unknown length
* 106548 108732: contig of 2185 bp in length
* 108733 108832: gap of unknown length
* 108833 111752: contig of 2920 bp in length
* 111753 111852: gap of unknown length
* 111853 113736: contig of 1884 bp in length
* 113737 113836: gap of unknown length
* 113837 118162: contig of 4326 bp in length
* 118163 118262: gap of unknown length
* 118263 121962: contig of 3700 bp in length
* 121963 122062: gap of unknown length
* 122063 124951: contig of 2889 bp in length
* 124952 125051: gap of unknown length
* 125052 128066: contig of 3015 bp in length
* 128067 128166: gap of unknown length
* 128167 131632: contig of 3466 bp in length
* 131633 131732: gap of unknown length
* 131733 135668: contig of 3936 bp in length
* 135669 135768: gap of unknown length
* 135769 139748: contig of 3980 bp in length
* 139749 139848: gap of unknown length
* 139849 142632: contig of 2784 bp in length
* 142633 142732: gap of unknown length
* 142733 145924: contig of 3192 bp in length
* 145925 146024: gap of unknown length

Query Match
Best Local Similarity 80.9%; Score 17.8; DB 2; Length 336873;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
DB 141009 TCCTTGGCCCGCCCTTGGGAT 140989

RESULT 47
LOCUS CQ662027 180 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 6953 from Patent WO02070737.
ACCESSION CQ662027
VERSION CQ662027.1 GI:42136652
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 6953 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
Location/Qualifiers
source
1 .180
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 79.1%; Score 17.4; DB 6; Length 180;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGG 19
DB 6 ATCCTTGCCCGCCCTTGG 24

RESULT 48
LOCUS CQ697578 262 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 42504 from Patent WO02070737.
ACCESSION CQ697578
VERSION CQ697578.1 GI:42249136
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 42504 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
Location/Qualifiers
source
1 .262
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 79.1%; Score 17.4; DB 6; Length 262;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGG 19
DB 98 ATCCTTGCCCGCCCTTGG 116

RESULT 49
LOCUS CQ664542 277 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 9468 from Patent WO02070737.
ACCESSION CQ664542
VERSION CQ664542.1 GI:42146052
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 9468 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
Location/Qualifiers
source
1 .277
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 79.1%; Score 17.4; DB 6; Length 277;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 ATCCTTGCCCCGCCCTTG 19
 |||||
 Db 116 ATCCTTGCCCCGCCCTTG 134

RESULT 50
 CQ398666/c
 LOCUS CQ398666 461 bp DNA linear PAT 23-JAN-2004
 DEFINITION Sequence 5737 from Patent WO0170979.

ACCESSION CQ398666
 VERSION CQ398666.1 GI:41306447

KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM

REFERENCE
 AUTHORS Lee,J. and Lillie,J.
 TITLE Genes, compositions, kits, and method for identification,
 assessment, prevention, and therapy of ovarian cancer
 JOURNAL Patent: WO 0170979-A 5737 27-SEP-2001;
 Millennium Pharmaceuticals, Inc. (US)

FEATURES
 source
 1...461
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 79.1%; Score 17.4; DB 6; Length 461;
 Best Local Similarity 94.7%; Pred.No. 2e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGCCCGCCCTTGAA 21
 |||||
 Db 133 CCTTGCCCGCCCTTGAA 115

Search completed: April 25, 2005, 12:18:41
 Job time : 501.276 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:06:08 ; Search time 219.431 Seconds
(without alignments)
593.509 Million cell updates/sec

Title: US-10-010-476-14
Perfect score: 22
Sequence: 1 ATCTTGCCCGCCCTTGAAT 22

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : N Geneseq_16Dec04:*

- 1: Geneseqn19808:*
- 2: Geneseqn19908:*
- 3: Geneseqn20008:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAx40729	Aax40729 GC-box d-
2	22	100.0	30	AAx40725	Aax40725 GC-box b-
3	22	100.0	30	AAx40724	Aax40724 GC-box a
4	22	100.0	50	AAx40728	Aax40728 GC-box c-
5	18.4	83.6	424	ACH04308	Ach04308 Human cDN
6	18.4	83.6	466	ABZ18450	Abz18450 Group III
7	17.8	80.9	540	ADL85957	Adl85957 DNA up-re
8	17.8	80.9	540	ADL85956	Adl85956 DNA up-re
9	17.8	80.9	994	ABZ52119	Abz52119 Aspergill
10	17.4	79.1	461	ADL38130	Adl38130 Human ova
11	17.4	79.1	461	ADI72995	Adi72995 Human ova
12	17.4	79.1	1789	ADR24724	Adr24724 Breast ca
13	17.2	78.2	818	ADR65453	Adr65453 Cotton cd
14	17	77.3	1051	ABN74461	Abn74461 Bovine em
15	16.8	76.4	347	ABN20125	Abn20125 Human ORF
16	16.8	76.4	596	AAx40726	Aax40726 Human rep
17	16.8	76.4	596	ABL98812	Abi98812 Human tes
18	16.8	76.4	618	AAx40726	Aax40726 Human rep
19	16.8	76.4	618	ABL98814	Abi98814 Human tes
20	16.8	76.4	621	AAx40726	Aax40726 Human rep

21	16.8	76.4	621	ABL98813	Abi98813 Human tes
22	16.8	76.4	716	ABI99929	Abi99929 Rat mucoc
23	16.8	76.4	787	ABI99930	Abi99930 Rat mucoc
24	16.8	76.4	1126	ABK49702	Abk49702 cDNA enco
25	16.8	76.4	2030	AAK94185	Aak94185 Human ful
26	16.8	76.4	2030	ADL30659	Adl30659 Full leng
27	16.8	76.4	2030	ADQ68003	Adq68003 Recombina
28	16.8	76.4	2513	ABL23652	Abi23652 Drosophil
29	16.8	76.4	4013	ABL20443	Abi20443 Drosophil
30	16.8	76.4	4098	ABL18683	Abi18683 Drosophil
31	16.8	76.4	7920	ABL18682	Abi18682 Drosophil
32	16.8	76.4	8557	ABL20442	Abi20442 Drosophil
33	16.8	76.4	50341	AAV22674	Aav22674 DNA seque
34	16.8	76.4	50341	AAZ39519	Aaz39519 L5 shuttl
35	16.8	76.4	52297	AAZ51411	Aaz51411 Mycobacte
36	16.8	76.4	52298	AAQ47357	Aaq47357 L5 mycoba
37	16.8	76.4	107330	ADQ97316	Adq97316 Mouse can
38	16.4	74.5	330	ABT05583	Abt05583 Mycobacte
39	16.4	74.5	903	ADJ44715	Adj44715 Plant cDN
40	16.4	74.5	1256	ABL29623	Abi29623 Drosophil
41	16.4	74.5	1567	ACA35376	Aca35376 Prokaryot
42	16.4	74.5	3549	ABL29622	Abi29622 Drosophil
43	16.4	74.5	4739	AAV17758	Aav17758 Mycobacte
44	16.4	74.5	32503	ABL12964	Abi12964 Drosophil
45	16.4	74.5	33454	ACN44588	Acn44588 Mouse gen
46	16.4	74.5	110000	AAI99682_01	AAI99682_01
47	16.4	74.5	110000	AAI99683_01	AAI99683_01
48	16.4	74.5	194534	ADQ97481	Adq97481 Human can
49	16.2	73.6	155	ACH91211	Ach91211 Human gen
50	16.2	73.6	222	ABK80858	Abk80858 Bacillus
51	16.2	73.6	257	ACH89088	Ach89088 Human gen
52	16.2	73.6	370	AA887085	AA887085 DNA encod
53	16.2	73.6	423	ABQ91356	Abq91356 M. capsul
54	16.2	73.6	445	AAAL00428	AAal00428 Human rep
55	16.2	73.6	445	ABA07504	Abao7504 Human ova
56	16.2	73.6	482	ABV13574	Abv13574 Human pro
57	16.2	73.6	484	AAK64074	AAk64074 Human imm
58	16.2	73.6	508	ACH77511	Ach77511 Human gen
59	16.2	73.6	545	ACH75360	Ach75360 Human gen
60	16.2	73.6	704	AAF65625	AAF65625 Novel hum
61	16.2	73.6	727	AAZ96452	Aaz96452 S. pneumo
62	16.2	73.6	762	ADS46365	Ads46365 Bacterial
63	16.2	73.6	837	ADT44430	Adt44430 Bacterial
64	16.2	73.6	1082	ABA14096	Abal4096 Human ner
65	16.2	73.6	1128	AAZ76704	Aaz76704 WSSV 664
66	16.2	73.6	1149	AAK63160	Aak63160 Human imm
67	16.2	73.6	1155	AAV72243	Aav72243 Drosophil
68	16.2	73.6	1155	ABL08709	Abi08709 Drosophil
69	16.2	73.6	1252	AAA94807	Aaa94807 Drosophil
70	16.2	73.6	1354	AAA72292	Aaa72292 Drosophil
71	16.2	73.6	1419	ADS56861	Ads56861 Bacterial
72	16.2	73.6	1500	AAI93722	Aai93722 Human pol
73	16.2	73.6	1787	ADB63520	Adb63520 Human cDN
74	16.2	73.6	2073	ADP10455	Adp10455 Reference
75	16.2	73.6	2149	ADM03180	Adm03180 Human cDN
76	16.2	73.6	2223	ADB63510	Adb63510 Human cDN
77	16.2	73.6	2244	ADQ23814	Adq23814 Human sof
78	16.2	73.6	2298	ADM02690	Adm02690 Human cDN
79	16.2	73.6	2299	RAZ54594	Raz54594 Human LIM
80	16.2	73.6	2368	AAD27808	Aad27808 Human pro
81	16.2	73.6	2634	AAAD39055	AAad39055 Human 577
82	16.2	73.6	2667	AAAD27807	AAad27807 Human pro
83	16.2	73.6	2828	AAAS87087	AAas87087 DNA encod
84	16.2	73.6	3234	ABL08708	Abi08708 Drosophil
85	16.2	73.6	3657	ACN92864	Acn92864 Breast ca
86	16.2	73.6	5998	ABZ24582	Abz24582 Human cel
87	16.2	73.6	6591	ADQ24342	Adq24342 Human sof
88	16.2	73.6	6852	ABL70311	Abi70311 Chemical
89	16.2	73.6	6852	AA61249	AA61249 Human gen
90	16.2	73.6	7008	ADB78645	Adb78645 Human ion
91	16.2	73.6	7008	ADB78639	Adb78639 Human ion
92	16.2	73.6	9699	ABA21144	Ab21144 Human ner
93	16.2	73.6	17457	ABA21143	Ab21143 Human ner

c 94	16.2	73.6	17458	5	ABA221142	Abaz1142 Human ner	167	15.8	71.8	94810	10	ADB72254	Adb72254 Human RAS
c 95	16.2	73.6	18234	4	AH62719	Ah62719 Shrimp wh	168	15.8	71.8	94810	10	ADE95764	Ade95764 Human RAS
c 96	16.2	73.6	18234	10	ABZ76687	Abz76687 White spo	c 169	15.8	71.8	96593	9	ADA02885	Ada02885 Mouse Blm
c 97	16.2	73.6	20137	2	AAZ27027	Aaz27027 Human CHD	c 170	15.8	71.8	96593	10	ADB72623	Adb72623 Mouse Blm
c 98	16.2	73.6	20138	2	AAZ26836	Aaz26836 Human CHD	c 171	15.8	71.8	96593	12	ADM74480	Adm74480 Murine ca
c 99	16.2	73.6	21596	4	AAK71582	Aak71582 Human lmm	c 172	15.8	71.8	96594	10	ADC85364	Adc85364 Human Pap
c 100	16.2	73.6	23071	2	AAZ27029	Aaz27029 Human CHD	c 173	15.8	71.8	110000	10	ACF67367	Acf67367 07
c 101	16.2	73.6	32816	6	ABK49701	Abk49701 Human tra	c 174	15.8	71.8	110000	10	ACF65384	1
c 102	16.2	73.6	32816	12	ADG16994	Adg16994 Human tra	c 175	15.8	71.8	160271	4	AAF85116	Aaf85116 Nucleotid
c 103	16.2	73.6	33916	13	ABD33626	Abd33626 Murine ca	c 176	15.8	71.8	160271	4	AAF85756	Aaf85756 Human chr
c 104	16.2	73.6	44118	10	ADL13708	Adl13708 Osteoarth	c 177	15.8	71.8	160271	4	AAF85750	Aaf85750 Bipolar a
c 105	16.2	73.6	70000	6	AAD42934	Aad42934 Human pho	c 178	15.8	71.8	160271	4	AAO66667	Aao66667 Human chr
c 106	16.2	73.6	90104	4	ABL12402	Ab112402 Drosophil	c 179	15.8	71.8	160271	4	AAO4864	Aao4864 Human chr
c 107	16.2	73.6	110000	13	ABD33780	1	c 180	15.8	71.8	160271	5	AAH23764	Aah23764 Human chr
c 108	16.2	73.6	139573	10	ADH58564	Adh58564 Human Na+	c 181	15.8	71.8	160271	5	AAH40997	Aah40997 160kb fra
c 109	16.2	73.6	156416	13	ABD32817	Abd32817 Human can	c 182	15.8	71.8	160271	5	AAO4858	Aao4858 Human chr
c 110	16.2	73.6	158001	12	ADL17884	Adl17884 Human pho	c 183	15.8	71.8	164991	10	ADL13635	Adl13635 Osteoarth
c 111	16.2	73.6	203132	13	ABD33364	Abd33364 Murine ca	c 184	15.8	71.8	272022	12	ADQ97126	Adq97126 Human can
c 112	16.2	73.6	305107	4	AH62689	Ah62689 Shrimp wh	c 185	15.6	70.9	80	12	ADM96117	Adm96117 Rat antis
c 113	16	72.7	255	10	ACD98140	Acd98140 Human col	c 186	15.6	70.9	93	6	ABV98337	Abv98337 Human pan
c 114	16	72.7	326	4	AA339325	Aas339325 Novel hum	c 187	15.6	70.9	178	9	ACH30721	Ach30721 Human bon
c 115	16	72.7	340	4	AA339329	Aas339329 Novel hum	c 188	15.6	70.9	296	6	ABL67346	Ab167346 Thyroid c
c 116	16	72.7	1359	5	AA65178	Aae65178 DNA encod	c 189	15.6	70.9	308	4	AAH71221	Aah71221 Human cer
c 117	16	72.7	1563	2	AAZ53007	Aaz53007 Human pro	c 190	15.6	70.9	322	6	ABV97178	Abv97178 Human pan
c 118	16	72.7	1597	5	AA65180	Aae65180 DNA encod	c 191	15.6	70.9	340	2	AAQ59910	Aaq59910 Human pan
c 119	16	72.7	1597	13	ACN40708	Acn40708 Tumour-as	c 192	15.6	70.9	368	6	ABV98029	Abv98029 Human pan
c 120	16	72.7	1597	13	ADR66182	Adr66182 Human pro	c 193	15.6	70.9	383	5	ABV17241	Abv17241 Human pro
c 121	16	72.7	1619	12	ADQ84817	Adq84817 Human tum	c 194	15.6	70.9	456	2	AAV86721	Aav86721 EST clone
c 122	16	72.7	1619	12	ADQ85892	Adq85892 Human tum	c 195	15.6	70.9	499	4	AAI20233	Aai20233 Human bre
c 123	16	72.7	1622	4	AH14860	Aah14860 Human cDN	c 196	15.6	70.9	524	5	ABV43084	Abv43084 Human pro
c 124	16	72.7	1655	13	ADR21626	Adr21626 Human enz	c 197	15.6	70.9	534	5	ABV47036	Abv47036 Human pro
c 125	16	72.7	13273	4	AA336849	Aas336849 Human car	c 198	15.6	70.9	600	12	ADL00134	Adl00134 Malignant
c 126	16	72.7	13273	10	AD647543	Ad647543 Human car	c 199	15.6	70.9	655	6	ABT09364	Abt09364 Phase-1 R
c 127	16	72.7	13273	13	ADJ08961	Adj08961 Human car	c 200	15.6	70.9	655	12	ADG45424	Adg45424 Liver inf
c 128	15.8	71.8	343	3	AA98195	Aas98195 Human ret	c 201	15.6	70.9	678	10	ADH83377	Adh83377 Enterococ
c 129	15.8	71.8	343	3	AA98196	Aas98196 Human ret	c 202	15.6	70.9	705	6	ABT09033	Abt09033 Phase-1 R
c 130	15.8	71.8	343	3	AA98191	Aas98191 Human ret	c 203	15.6	70.9	705	12	ADG45731	Adg45731 Liver inf
c 131	15.8	71.8	343	3	AA98199	Aas98199 Human ret	c 204	15.6	70.9	782	6	ABQ51980	Abq51980 Oligonucl
c 132	15.8	71.8	343	3	AA98190	Aas98190 Human ret	c 205	15.6	70.9	782	6	ABQ51981	Abq51981 Oligonucl
c 133	15.8	71.8	453	4	ABR26612	Abra26612 Probe #50	c 206	15.6	70.9	981	6	ABN74439	Abn74439 Bovine em
c 134	15.8	71.8	453	4	AAK05068	Aak05068 Human bra	c 207	15.6	70.9	1177	2	AAV13960	Aav13960 Rat inter
c 135	15.8	71.8	453	6	ABSO5313	Abso5313 Human gen	c 208	15.6	70.9	1177	2	AAV13959	Aav13959 Rat inter
c 136	15.8	71.8	472	4	AAI14594	Aai14594 Probe #45	c 209	15.6	70.9	1249	10	ADB52637	Adb52637 Primary r
c 137	15.8	71.8	472	4	ABA56322	Abas6322 Human foe	c 210	15.6	70.9	1683	2	AAT86246	Aat86246 cDNA enco
c 138	15.8	71.8	472	4	AAI35964	Aai35964 Probe #46	c 211	15.6	70.9	1693	12	ADQ24476	Adq24476 Human sof
c 139	15.8	71.8	472	4	ABR45808	Abra45808 Human bre	c 212	15.6	70.9	1729	13	ADT44742	Adt44742 Bacterial
c 140	15.8	71.8	472	4	ABA25962	Abas25962 Probe #44	c 213	15.6	70.9	1966	3	AACT78024	Aac78024 Human can
c 141	15.8	71.8	472	4	AAK30001	Aak30001 Human bon	c 214	15.6	70.9	2000	3	AACT10825	Aac10825 Human sec
c 142	15.8	71.8	472	4	AAK04501	Aak04501 Human bra	c 215	15.6	70.9	2000	6	ABZ17505	Abz17505 Arabidops
c 143	15.8	71.8	472	5	AAI04406	Aai04406 Probe #43	c 216	15.6	70.9	2000	6	ABZ17571	Abz17571 Arabidops
c 144	15.8	71.8	472	6	ABS04574	Abso4574 Human gen	c 217	15.6	70.9	2033	4	AAH73269	Aah73269 Human cer
c 145	15.8	71.8	734	12	ADJ43268	Adj43268 Plant cDN	c 218	15.6	70.9	2115	3	AACT10826	Aac10826 Human sec
c 146	15.8	71.8	1008	6	ABN74608	Abn74608 Bovine em	c 219	15.6	70.9	2205	12	ADN17324	Adn17324 H. annosu
c 147	15.8	71.8	1038	10	ACF68026	Acf68026 Phototrab	c 220	15.6	70.9	2331	4	ABL09127	Ab109127 Drosophil
c 148	15.8	71.8	1115	5	AA76196	Aas76196 DNA encod	c 221	15.6	70.9	2364	4	AAV41207	Aav41207 cDNA enco
c 149	15.8	71.8	1299	8	ACA25973	Acas25973 Prokaryot	c 222	15.6	70.9	2558	5	ABV28386	Abv28386 Human pro
c 150	15.8	71.8	1723	2	AAV44296	Aav44296 Human sec	c 223	15.6	70.9	2677	2	AAV31209	Aav31209 E. coli J
c 151	15.8	71.8	1723	5	AA984870	Aa984870 Human cDN	c 224	15.6	70.9	2686	5	ADL63328	Adl63328 Human ova
c 152	15.8	71.8	2181	6	ABN95812	Abn95812 Gene #231	c 225	15.6	70.9	2844	11	ACN91342	Acn91342 Breast ca
c 153	15.8	71.8	2330	4	ABL26954	Ab126954 Drosophil	c 226	15.6	70.9	2992	12	ADP10494	Adp10494 Reference
c 154	15.8	71.8	2340	4	ABL11005	Ab111005 Drosophil	c 227	15.6	70.9	3049	6	ABN59759	Abn59759 Novel hum
c 155	15.8	71.8	2349	10	ADB62994	Abd62994 Human cDN	c 228	15.6	70.9	3113	12	ADJ74953	Adj74953 Marker ge
c 156	15.8	71.8	2573	10	ABT42316	Abt42316 Toxicity	c 229	15.6	70.9	3305	4	AAI97924	Aai97924 Human neu
c 157	15.8	71.8	3003	8	ACC46294	Acc46294 Human dit	c 230	15.6	70.9	3305	4	AAI98080	Aai98080 Human neu
c 158	15.8	71.8	3263	8	ACC46303	Acc46303 Human dit	c 231	15.6	70.9	3490	12	ADP84460	Adp84460 Human bre
c 159	15.8	71.8	3665	4	ABL11522	Ab111522 Drosophil	c 232	15.6	70.9	3719	2	AAQ55142	Aaq55142 Enterococ
c 160	15.8	71.8	3702	8	ADA71172	Ada71172 Rice gene	c 233	15.6	70.9	4195	4	AAH98347	Aah98347 Human EST
c 161	15.8	71.8	5592	4	ABL11004	Ab111004 Drosophil	c 234	15.6	70.9	4315	2	AAI12938	Aai12938 Enterococ
c 162	15.8	71.8	10102	10	ADC68678	Adc68678 Human GPC	c 235	15.6	70.9	4315	2	ABS98733	Abs98733 Enterococ
c 163	15.8	71.8	13175	4	ABL02816	Ab102816 Drosophil	c 236	15.6	70.9	4331	4	ABL09126	Ab109126 Drosophil
c 164	15.8	71.8	23899	4	ABL0362	Ab10362 Drosophil	c 237	15.6	70.9	4355	8	ABX63010	Abx63010 Human cDN
c 165	15.8	71.8	52637	13	ADR67083	Adr67083 Mouse can	c 238	15.6	70.9	4385	4	AA522674	Aa522674 Human cDN
c 166	15.8	71.8	94810	9	ADA02516	Ada02516 Human RAS	c 239	15.6	70.9	4562	13	ADR15041	Adr15041 Rat elect

[illegible]

386	15.2	69.1	379	5	ABV56389	Human pro	Abv56389	Human pro	10	ADG10444	Human STA
387	15.2	69.1	396	8	ABX65117	Human gen	Aat65117	Human gen	2	ABK10951	DNA encod
388	15.2	69.1	397	2	AA773660	ICAM-2 ge	Aat73660	ICAM-2 ge	2	ABL08529	Drosophil
389	15.2	69.1	414	6	ABN68206	Streptoco	Abn68206	Streptoco	13	ADP56376	Human PRO
390	15.2	69.1	416	4	AAK73524	Human imm	Aak73524	Human imm	5	AA83450	DNA encod
391	15.2	69.1	416	4	AAK73523	Human imm	Aak73523	Human imm	3	AA874646	Human ORF
392	15.2	69.1	422	13	ADQ49534	Novel can	Adq49534	Novel can	11	ABD03096	Pseudomon
393	15.2	69.1	455	9	ACL21677	DNA clone	Ac121677	DNA clone	11	ADM03345	Human cDN
394	15.2	69.1	458	9	ACL21678	DNA clone	Ac121678	DNA clone	11	ADM03345	Human cDN
395	15.2	69.1	476	9	ACL21679	DNA clone	Ac121679	DNA clone	10	ADG10446	Human STA
396	15.2	69.1	482	9	ACL21671	DNA clone	Ac121671	DNA clone	10	ABL13475	Drosophil
397	15.2	69.1	483	6	ABN95835	Gene #233	Abn95835	Gene #233	5	AA871055	DNA encod
398	15.2	69.1	485	9	ACL21680	DNA clone	Ac121680	DNA clone	13	ADR84662	Aspergill
399	15.2	69.1	485	9	ACL21676	DNA clone	Ac121676	DNA clone	10	ADD46886	Rat gene
400	15.2	69.1	486	9	ACL21683	DNA clone	Ac121683	DNA clone	10	ADD46882	Rat gene
401	15.2	69.1	501	9	ACH37427	Human end	Ach37427	Human end	4	ABL12855	Drosophil
402	15.2	69.1	501	9	ACL21669	DNA clone	Ac121669	DNA clone	3	AA818287	Lung canc
403	15.2	69.1	501	9	ACL21672	DNA clone	Ac121672	DNA clone	3	ADR06903	Bacterial
404	15.2	69.1	503	5	ADL38018	Human ova	Adl38018	Human ova	13	ADS45801	Bacterial
405	15.2	69.1	503	5	ADL38018	Human ova	Adl38018	Human ova	13	AA823193	DNA encod
406	15.2	69.1	508	9	ACL21681	DNA clone	Ac121681	DNA clone	4	AA809264	Human DES
407	15.2	69.1	534	6	ABQ29349	Oligonuc1	Abq29349	Oligonuc1	4	AAK53152	Human pol
408	15.2	69.1	534	6	ABQ29348	Oligonuc1	Abq29348	Oligonuc1	13	ACN40789	Tumour-as
409	15.2	69.1	534	6	ABQ17394	Oligonuc1	Abq17394	Oligonuc1	13	ADR06903	Full leng
410	15.2	69.1	534	6	ABQ17395	Oligonuc1	Abq17395	Oligonuc1	4	AA823106	DNA encod
411	15.2	69.1	561	8	AB253373	Aspergill	Ab253373	Aspergill	4	AAK52168	Human pol
412	15.2	69.1	566	12	ACH89728	Human gen	Ach89728	Human gen	6	ABQ54830	Human ova
413	15.2	69.1	578	9	ACL21674	DNA clone	Ac121674	DNA clone	3	ABT17915	Aspergill
414	15.2	69.1	578	12	ACH80054	Human gen	Ach80054	Human gen	8	ABT19729	Aspergill
415	15.2	69.1	580	9	ACL21685	DNA clone	Ac121685	DNA clone	3	ACN44319	Human mRN
416	15.2	69.1	634	13	ADR60120	Cotton cD	Adr60120	Cotton cD	11	ACN44319	Human mRN
417	15.2	69.1	649	9	ACL19921	DNA clone	Ac119921	DNA clone	13	AAH17723	Human nar
418	15.2	69.1	654	9	ACL21682	DNA clone	Ac121682	DNA clone	5	AAH78735	Human HAJ
419	15.2	69.1	669	9	AB271799	Human can	Ab271799	Human can	4	AAH46858	Human ser
420	15.2	69.1	673	9	ACL21684	DNA clone	Ac121684	DNA clone	5	AAH46858	Human ser
421	15.2	69.1	675	10	ADD33429	Mouse mit	Add33429	Mouse mit	4	AD41923	Human nov
422	15.2	69.1	679	3	AA772319	Human Rec	Aa772319	Human Rec	10	ABQ83667	Human MDD
423	15.2	69.1	680	9	ACL19917	DNA clone	Ac119917	DNA clone	4	ABQ83667	Human MDD
424	15.2	69.1	712	2	AAQ71011	lyan-1 ge	Aaq71011	lyan-1 ge	4	AAK83412	Human imm
425	15.2	69.1	740	3	AAK02334	Human col	Aak02334	Human col	13	ADR07984	Full leng
426	15.2	69.1	792	13	ADT46056	Bacterial	Adt46056	Bacterial	4	ABL08528	Drosophil
427	15.2	69.1	810	2	AAQ22246	Sequence	Aaq22246	Sequence	4	ABL12392	Drosophil
428	15.2	69.1	846	8	AAK48490	Wheat DNA	Aak48490	Wheat DNA	4	ABX63767	Human cDN
429	15.2	69.1	846	12	ADO62473	Transcrip	Ado62473	Transcrip	8	AB128488	Drosophil
430	15.2	69.1	935	6	ABQ22588	Oligonuc1	Abq22588	Oligonuc1	2	AAQ31768	Genomic H
431	15.2	69.1	935	6	ABQ22589	Oligonuc1	Abq22589	Oligonuc1	2		
432	15.2	69.1	1003	6	ABN74036	Bovine em	Abn74036	Bovine em	4		
433	15.2	69.1	1035	4	AAQ65409	cDNA enco	Aaq65409	cDNA enco	4		
434	15.2	69.1	1123	2	AAQ65409	cDNA enco	Aaq65409	cDNA enco	4		
435	15.2	69.1	1124	2	AAQ73596	Lolium so	Aaq73596	Lolium so	4		
436	15.2	69.1	1189	12	ADQ24500	Human sof	Adq24500	Human sof	4		
437	15.2	69.1	1274	13	ADS60746	Bacterial	Ads60746	Bacterial	4		
438	15.2	69.1	1319	3	AAQ70537	Fusarium	Aaq70537	Fusarium	4		
439	15.2	69.1	1389	8	ACA45627	Prokaryot	Aca45627	Prokaryot	4		
440	15.2	69.1	1403	4	AAK65360	Human imm	Aak65360	Human imm	4		
441	15.2	69.1	1403	4	AAK65360	Human imm	Aak65360	Human imm	4		
442	15.2	69.1	1403	4	AAK65361	Human imm	Aak65361	Human imm	4		
443	15.2	69.1	1403	4	AAK65361	Human imm	Aak65361	Human imm	4		
444	15.2	69.1	1452	10	ACC61055	Gene sequ	Acc61055	Gene sequ	4		
445	15.2	69.1	1452	10	ADK62761	Disease t	Adk62761	Disease t	4		
446	15.2	69.1	1470	12	ADN43056	Human sec	Adn43056	Human sec	4		
447	15.2	69.1	1509	11	ABD02998	Pseudomon	Abd02998	Pseudomon	4		
448	15.2	69.1	1527	4	ABL10329	Drosophil	Ab110329	Drosophil	4		
449	15.2	69.1	1542	10	ABD53646	Primary r	Abd53646	Primary r	4		
450	15.2	69.1	1568	6	ABK28657	Human cDN	Abk28657	Human cDN	4		
451	15.2	69.1	1572	13	ABD02816	Pseudomon	Abd02816	Pseudomon	4		
452	15.2	69.1	1591	13	ADS54848	Bacterial	Ads54848	Bacterial	4		
453	15.2	69.1	1601	4	AAH98368	Arabidops	Aah98368	Arabidops	4		
454	15.2	69.1	1613	2	AAK26196	DNA sequ	Aak26196	DNA sequ	4		
455	15.2	69.1	1718	3	AAK76916	Human ORF	Aac76916	Human ORF	4		
456	15.2	69.1	1746	4	ABU12393	Drosophil	Ab112393	Drosophil	4		
457	15.2	69.1	1952	4	AAH29727	S cerevis	Aah29727	S cerevis	4		
458	15.2	69.1	1962	6	ABN79808	Fungal ZB	Abn79808	Fungal ZB	4		

ALIGNMENTS

RESULT 1

AAK40729
ID AAK40729 standard; DNA; 22 BP.

XX AC AAK40729;
XX AC AAK40729;
XX DT 16-JUN-1999 (first entry)
XX GC-box d-MET nucleotide sequence.

KW DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
ss.

OS Synthetic.

XX WO9912027-A1.

XX 11-MAR-1999.

XX 12-JUN-1998; 98WO-US012351.

XX 29-AUG-1997; 97US-005741P.

XX (REGC) UNIV CALIFORNIA.

PA

XX PI Reich NO, Flynn J;
XX WPI; 1999-205256/17.
XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
PT used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
XX Claim 10; Page 63; 114pp; English.
XX The specification describes a synthetic oligonucleotide comprising a C-5
CC methylcytosine which recognizes and binds an allosteric site on DNA
CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
CC activity associated with the allosteric site. The synthetic inhibitor can
CC be used to inhibit methylation of DNA. It can also be used to inhibit
CC proliferation of cancer cells. The inhibitor can also be used to treat a
CC disorder of development, which is linked to a genetic locus regulated by
CC methylation, such as Huntington's disease, Down's syndrome, and disorders
CC associated with a Hox gene. The present sequence appears in the
CC specification
XX
XX Sequence 22 BP; 3 A; 10 C; 3 G; 6 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 22; DB 2; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 2.4;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATCCTTGCCCGCCCGCCCTTGAAT 22
XX DB 1 ATCCTTGCCCGCCCGCCCTTGAAT 22
XX
XX RESULT 2
XX AAX40725
XX ID AAX40725 standard; DNA; 30 BP.
XX AC AAX40725;
XX DT 16-JUN-1999 (first entry)
XX DE GC-box b-MET nucleotide sequence.
XX
XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW ss.
XX
XX Synthetic.
XX
XX WO9912027-A1.
XX
XX AC AAX40725;
XX DT 16-JUN-1999 (first entry)
XX DE GC-box b-MET nucleotide sequence.
XX
XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW ss.
XX
XX Synthetic.
XX
XX WO9912027-A1.
XX
XX PD 11-MAR-1999.
XX
XX PF 12-JUN-1999; 98WO-US012351.
XX
XX PR 29-AUG-1997; 97US-0057411P.
XX
XX PA (REGC) UNIV CALIFORNIA.
XX
XX PI Reich NO, Flynn J;
XX WPI; 1999-205256/17.
XX
XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
PT used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
XX Claim 10; Page 39; 114pp; English.
XX The specification describes a synthetic oligonucleotide comprising a C-5
CC methylcytosine which recognizes and binds an allosteric site on DNA
CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
CC activity associated with the allosteric site. The synthetic inhibitor can
CC be used to inhibit methylation of DNA. It can also be used to inhibit

CC proliferation of cancer cells. The inhibitor can also be used to treat a
CC disorder of development, which is linked to a genetic locus regulated by
CC methylation, such as Huntington's disease, Down's syndrome, and disorders
CC associated with a Hox gene. The present sequence appears in the
CC specification
XX
XX Sequence 30 BP; 3 A; 14 C; 5 G; 8 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 22; DB 2; Length 30;
XX Best Local Similarity 100.0%; Pred. No. 2.4;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATCCTTGCCCGCCCGCCCTTGAAT 22
XX DB 5 ATCCTTGCCCGCCCGCCCTTGAAT 26
XX
XX RESULT 3
XX AAX40724/c
XX ID AAX40724 standard; DNA; 30 BP.
XX AC AAX40724;
XX DT 16-JUN-1999 (first entry)
XX DE GC-box a nucleotide sequence of the specification.
XX
XX DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW ss.
XX
XX Synthetic.
XX
XX WO9912027-A1.
XX
XX PD 11-MAR-1999.
XX
XX PF 12-JUN-1999; 98WO-US012351.
XX
XX PR 29-AUG-1997; 97US-0057411P.
XX
XX PA (REGC) UNIV CALIFORNIA.
XX
XX PI Reich NO, Flynn J;
XX WPI; 1999-205256/17.
XX
XX DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
PT used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
XX Example 2; Page 39; 114pp; English.
XX The specification describes a synthetic oligonucleotide comprising a C-5
CC methylcytosine which recognizes and binds an allosteric site on DNA
CC cytosine methyltransferase (DNMTase); thereby modulating DNMTase
CC activity associated with the allosteric site. The synthetic inhibitor can
CC be used to inhibit methylation of DNA. It can also be used to inhibit
CC proliferation of cancer cells. The inhibitor can also be used to treat a
CC disorder of development, which is linked to a genetic locus regulated by
CC methylation, such as Huntington's disease, Down's syndrome, and disorders
CC associated with a Hox gene. The present sequence appears in the
CC specification
XX
XX Sequence 30 BP; 8 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 22; DB 2; Length 30;
XX Best Local Similarity 100.0%; Pred. No. 2.4;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATCCTTGCCCGCCCGCCCTTGAAT 22
XX DB 26 ATCCTTGCCCGCCCGCCCTTGAAT 5

RESULT 4
 AAX40728
 ID AAX40728 standard; DNA; 50 BP.
 XX
 AC AAX40728;
 XX
 DT 16-JUN-1999 (first entry)
 XX
 DE GC-box c-MET nucleotide sequence.
 XX
 KW DNA cytosine methyltransferase; DCMase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 KW ss.
 OS Synthetic.
 XX
 PN W09912027-A1.
 XX
 PD 11-MAR-1999.
 XX
 PF 12-JUN-1998; 98WO-US012351.
 XX
 PR 29-AUG-1997; 97US-0057411P.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Reich NO. Flynn J;
 XX
 DR WPI; 1999-205256/17.
 XX
 PT DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
 PT used to inhibit methylation of DNA, and proliferation of cancer cells.
 XX
 PS Claim 10; Page 63; 114pp; English.
 CC The specification describes a synthetic oligonucleotide comprising a C-5
 CC methylcytosine which recognizes and binds an allosteric site on DNA
 CC cytosine methyltransferase (DCMase); thereby modulating DCMase
 CC activity associated with the allosteric site. The synthetic inhibitor can
 CC be used to inhibit methylation of DNA. It can also be used to inhibit
 CC proliferation of cancer cells. The inhibitor can also be used to treat a
 CC disorder of development, which is linked to a genetic locus regulated by
 CC methylation, such as Huntington's disease, Down's syndrome, and disorders
 CC associated with a Hox gene. The present sequence appears in the
 CC specification
 XX
 SQ Sequence 50 BP; 8 A; 27 C; 5 G; 10 T; 0 U; 0 Other;
 Query Match 100.0%; Score 22; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATCCTTGCCCGCCCTTGAAT 22
 DB 15 ATCCTTGCCCGCCCTTGAAT 36
 RESULT 5
 ACH04308/c
 ID ACH04308 standard; cDNA; 424 BP.
 XX
 AC ACH04308;
 XX
 DT 26-SEP-2003 (first entry)
 XX
 DE Human cDNA differentially expressed in lung cancer #513.
 XX
 KW Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
 KW respiratory disorder; lung cancer; asthma; human.
 XX

OS Homo sapiens.
 XX
 PN US2003065157-A1.
 XX
 PD 03-APR-2003.
 XX
 PF 04-APR-2002; 2002US-00116802.
 XX
 PR 04-APR-2001; 2001US-0281593P.
 XX
 PA (LASE/) LASEK A W.
 XX
 PI Lasek AW;
 XX
 DR WPI; 2003-540803/51.
 XX
 PT New combination comprising cDNAs that are differentially expressed in
 PT respiratory disorders, useful for diagnosing or treating respiratory
 PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
 PT emphysema or asthma.
 XX
 PS Claim 1; Page; 39pp; English.
 XX
 CC The invention relates to a combination comprising cDNAs or their
 CC complements that are differentially expressed in respiratory disorder.
 CC The combination is useful for preparing a composition for diagnosing or
 CC treating respiratory disorders e.g. lung cancer, chronic obstructive
 CC pulmonary disease, emphysema or asthma. The present sequence represents
 CC human cDNA differentially expressed during lung cancer
 XX
 SQ Sequence 424 BP; 73 A; 149 C; 132 G; 70 T; 0 U; 0 Other;
 Query Match 83.6%; Score 18.4; DB 9; Length 424;
 Best Local Similarity 95.0%; Pred. No. 1.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 CCTTGCCCGCCCTTGAAT 22
 DB 343 CCGTGCCCGCCCTTGAAT 324
 RESULT 6
 ABZ18450/c
 ID ABZ18450 standard; cDNA; 466 BP.
 XX
 AC ABZ18450;
 XX
 DT 23-JAN-2003 (first entry)
 XX
 DE Group III cDNA cancer related clone SEQ ID NO:876.
 XX
 KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
 KW immune response; virology; immunology; microbiology; molecular biology;
 KW recombinant DNA technology; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200278516-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US010421.
 XX
 PR 30-MAR-2001; 2001US-0280255P.
 PR 28-AUG-2001; 2001US-0315563P.
 PR 09-JAN-2002; 2002US-0347313P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Wang S, Bangur CS, Gaiger A;
 XX
 DR WPI; 2003-058387/05.
 XX

PT New immunogenic polynucleotides or polypeptides useful for diagnosing, preventing and treating cancer expressing CT or CP mRNA antigens, and in
PT virology, immunology, microbiology, molecular biology and recombinant DNA
PT techniques.

XX Claim 1; SEQ ID NO 876; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
CC ABP54446 to ABP54472 represent protein (II) sequences, from the present
CC invention. (I) and (II) have cytosolic activity and can be used in gene
CC therapy and vaccines. (I), (II), antibodies and compositions from the
CC present invention are useful for diagnosing, preventing and treating
CC cancer, which expresses CT or CP mRNA antigens. They are useful for
CC stimulating immune response. They can also be useful in virology,
CC immunology, microbiology, molecular biology and recombinant DNA
CC techniques. N.B. The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pt_sequences

XX SQ Sequence 466 BP; 78 A; 165 C; 144 G; 79 T; 0 U; 0 Other;
Query Match 83.6%; Score 18.4; DB 8; Length 466;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGCCCGCCGCTTGAAT 22
DB 343 CCGTGCCCGCCGCTTGAAT 324
|||||

RESULT 7
ADL85957
ID ADL85957 standard; DNA; 540 BP.
XX AC ADL85957;
XX DT 20-MAY-2004 (first entry)
XX DE DNA up-regulated in murine multipotent progenitor cells SeqID 2350.
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX OS Mus sp.
XX PN WO2003093445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US014114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PS (STOW-) STOWERS INST MEDICAL RES.
XX PA Li L;
XX PI WPI; 2004-022656/02.
XX DR Classifying an unknown multi-lineage affiliated gene comprises isolating
XX expressed nucleic acid sequences from the discrete cell sub-populations.
XX Claim 8; SEQ ID NO 2350; 123pp; English.

XX This invention relates to a novel method for predicting gene potential by
CC associating nucleic acid sequences of unknown function with particular
CC sub-population profiles. Specifically, it refers to classifying an
CC unknown multi-lineage affiliated gene by collecting hybridisation data to
CC develop a gene expression map, in order to determine the discrete sub-
CC population where it is expressed. The present invention describes methods
CC for predicting the lineage commitment of genes associated with the self-
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC

CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
CC referred to as bone marrow stem cells populations. As such, these methods
CC can be used to identify associated multi-lineage affiliated genes and
CC hence the underlying molecular mechanisms in physiological haematopoietic
CC development. This polynucleotide sequence is DNA associated with a murine
CC MPP sub population of cells of the invention.

XX SQ Sequence 540 BP; 93 A; 146 C; 157 G; 128 T; 0 U; 16 Other;
Query Match 80.9%; Score 17.8; DB 12; Length 540;
Best Local Similarity 90.5%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCGCTTGAAT 22
DB 280 TCCTTGCCCGCCGCTTGAAT 300
|||||

RESULT 8
ADL85956
ID ADL85956 standard; DNA; 540 BP.
XX AC ADL85956;
XX DT 20-MAY-2004 (first entry)
XX DE DNA up-regulated in murine multipotent progenitor cells SeqID 2349.
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX OS Mus sp.
XX PN WO2003093445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US014114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PS (STOW-) STOWERS INST MEDICAL RES.
XX PA Li L;
XX PI WPI; 2004-022656/02.
XX DR Classifying an unknown multi-lineage affiliated gene comprises isolating
XX expressed nucleic acid sequences from the discrete cell sub-populations.
XX Claim 8; SEQ ID NO 2349; 123pp; English.

XX This invention relates to a novel method for predicting gene potential by
CC associating nucleic acid sequences of unknown function with particular
CC sub-population profiles. Specifically, it refers to classifying an
CC unknown multi-lineage affiliated gene by collecting hybridisation data to
CC develop a gene expression map, in order to determine the discrete sub-
CC population where it is expressed. The present invention describes methods
CC for predicting the lineage commitment of genes associated with the self-
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
CC referred to as bone marrow stem cells populations. As such, these methods
CC can be used to identify associated multi-lineage affiliated genes and
CC hence the underlying molecular mechanisms in physiological haematopoietic
CC development. This polynucleotide sequence is DNA associated with a murine
CC MPP sub population of cells of the invention.

XX SQ Sequence 540 BP; 93 A; 146 C; 157 G; 128 T; 0 U; 16 Other;
Query Match 80.9%; Score 17.8; DB 12; Length 540;

```
Best Local Similarity 90.5%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 280 TCCTTGCCCGCCCTTGAAT 300

RESULT 9
ABZ52119
ID ABZ52119 standard; cDNA; 994 BP.
XX AC
XX ABZ52119;
XX DT 28-MAR-2003 (first entry)
XX DE Aspergillus oryzae polynucleotide SEQ ID NO 1232.
XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX OS expressed sequence tag; gene; ss.
XX AS Aspergillus oryzae.
XX PN WO200279476-A1.
XX PD 10-OCT-2002.
XX PF 22-MAR-2002; 2002WO-IB000890.
XX PR 30-MAR-2001; 2001JP-00098371.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (NARE-) NAT RES INST BREWING.
XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX DR WPI; 2003-046817/04.
XX PT Detection of expression of specific Aspergillus genes for monitoring the
XX PT fermentation and growth conditions of the fungus, using DNA probes.
XX PS Claim 1; SEQ ID NO 1232; 48pp + Sequence Listing; Japanese.
XX CC The invention relates to a polynucleotide having any of 6006 specific
XX CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
XX CC specific culture conditions including one or more of eutrophic,
XX CC oligotrophic, solid, early germination, alkaline, high temperature, low
XX CC temperature or maltose culture or polynucleotides stringently hybridising
XX CC to these sequences. The polynucleotides are useful for monitoring the
XX CC progress of fermentation and the growth conditions of a fungus,
XX CC especially of Aspergillus oryzae which is widely used in industrial
XX CC fermentation. Also monitoring for fungal contamination. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 994 BP; 261 A; 296 C; 260 G; 176 T; 0 U; 1 Other;

Query Match 80.9%; Score 17.8; DB 8; Length 994;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 591 TCCTTGCCCGCCCTTGAAT 611

RESULT 10
ADL38130/c
ID ADL38130 standard; DNA; 461 BP.
XX
```

```
AC ADL38130;
XX 20-MAY-2004 (first entry)
XX Human ovarian cancer DNA marker #12020.
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX OS Homo sapiens.
XX PN WO200170979-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US009126.
XX PR 21-MAR-2000; 2000US-0191031P.
XX PR 25-MAY-2000; 2000US-0207124P.
XX PR 15-JUN-2000; 2000US-0211940P.
XX PR 07-JUL-2000; 2000US-0216820P.
XX PR 25-JUL-2000; 2000US-0220661P.
XX PR 21-DEC-2000; 2000US-0257672P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lee J, Lillie J;
XX DR WPI; 2001-611502/70.
XX PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX PT cancer cells as compared to their normal non-cancerous ovarian cells are
XX PT used to characterize stage, grade, histological type of ovarian cancer.
XX PS Disclosure; SEQ ID NO 12020; 106pp; English.
XX CC The invention relates to nucleic acid markers which are overexpressed in
XX CC ovarian cancer cells as compared to their expression in normal (i.e. non-
XX CC cancerous) ovarian cells. The invention also relates to polypeptides
XX CC encoded by the markers, antibodies that selectively bind to the
XX CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX CC of developing ovarian cancer involving inhibiting expression of a gene
XX CC corresponding to a marker of the invention and a method of treating a
XX CC patient afflicted with ovarian cancer comprising providing to cells of
XX CC the patient an antisense oligonucleotide complementary to a marker of the
XX CC invention. The markers are useful for assessing if a patient is afflicted
XX CC with ovarian cancer, which involves comparing the level of expression of
XX CC a marker in a patient sample and a normal level of expression of the
XX CC marker in a control non-ovarian cancer sample. A difference between the
XX CC expression levels indicates ovarian cancer. The level of expression of a
XX CC marker corresponds to a secreted protein or to a transcribed
XX CC polynucleotide or its portion. The level of expression of the marker is
XX CC assessed by detecting the presence in the sample, a protein or protein
XX CC fragment corresponding to the marker. The presence of protein or protein
XX CC fragment is detected using an antibody that specifically binds with the
XX CC protein or protein fragment. Alternatively, the level of expression of
XX CC the marker is assessed by detecting the presence of a transcribed
XX CC polynucleotide which anneals with the marker or anneals with a portion of
XX CC the polynucleotide comprising the marker, under stringent conditions. The
XX CC marker is also used for monitoring the progression of ovarian cancer in a
XX CC patient which involves detecting expression of the marker in a patient
XX CC sample at a first point in time, repeating the method at a subsequent
XX CC time and comparing the level of expression. The method is carried out
XX CC using an ovarian tissue sample. A composition comprising a marker,
XX CC polypeptide or antibody of the invention is used to treat ovarian cancer.
XX CC This sequence represents a human ovarian cancer DNA marker of the
XX CC invention.
XX SQ Sequence 461 BP; 109 A; 80 C; 94 G; 139 T; 0 U; 39 Other;

Query Match 79.1%; Score 17.4; DB 5; Length 461;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 3 CCTTGGCCCGCCCTTGAA 21
|||||
Db 133 CCTTCCCGCCCGCCCTTGAA 115

RESULT 11
ADI72995/C
ID ADI72995.standard; DNA; 461 BP.
AC ADI72995;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human ovarian cancer DNA marker #5737.
XX
XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
KW
XX Homo sapiens.
OS
XX WO200170979-A2.
PN
XX 27-SEP-2001.
PD
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
XX 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX Lee J, Lillie J;
PI
XX WPI; 2001-611502/70.
DR
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.
PT
XX
XX Disclosure; SEQ ID NO 5737; 106pp; English.

CC The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer.

CC This sequence represents a human ovarian cancer DNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC
XX
SQ Sequence 461 BP; 109 A; 80 C; 94 G; 139 T; 0 U; 39 Other;
Query Match 79.1%; Score 17.4; DB 5; Length 461;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAA 21
|||||
Db 133 CCTTCCCGCCCGCCCTTGAA 115

RESULT 12
ADR24724
ID ADR24724 standard; DNA; 1789 BP.
XX
AC ADR24724;
XX
XX 21-OCT-2004 (first entry)
DT
XX Breast cancer prognosis marker #585.
DE
XX ds; breast cancer; prognosis; gene expression; diagnosis.
KW
XX Homo sapiens.
OS
XX WO2004065545-A2.
XX
XX 05-AUG-2004.
PD
XX 15-JAN-2004; 2004WO-US001100.
PF
XX 15-JAN-2003; 2003US-00342887.
PR
XX (ROSE-) ROSETTA INPHARMATICS LLC.
PA (NECA-) NETHERLANDS CANCER INST.
XX
XX Van't Veer LJ, He Y;
PI
XX WPI; 2004-593473/57.

CC Classifying a breast cancer patient according to prognosis comprises determining the similarity between the level of expression of each of five genes in a cell sample taken from patient, to control levels.

XX
XX Disclosure; SEQ ID NO 585; 226pp; English.

CC The invention relates to a method of classifying a breast cancer patient according to prognosis by determining the similarity between the level of expression of each of five genes for which markers are listed in the specification, in a cell sample taken from the breast cancer patient, to control levels of expression for each respective five genes to obtain a patient similarity value. The methods are useful for classifying a breast cancer patient according to prognosis. Kits and computer program products are useful for data analysis using the diagnostic, prognostic and statistical methods of the invention. This sequence corresponds to a marker used in the method of the invention.

XX
SQ Sequence 1789 BP; 355 A; 655 C; 437 G; 342 T; 0 U; 0 Other;
Query Match 79.1%; Score 17.4; DB 13; Length 1789;
Best Local Similarity 94.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTG 19
|||||
Db 1622 ATCCTTGCCCGCCCTTG 1640

RESULT 13
AD65453
ID AD65453 standard; cDNA; 818 BP.
XX
AC AD65453;
XX
DT 02-DEC-2004 (first entry)
XX
DE Cotton cDNA sequence, SEQ ID 6234.
XX
KW Cotton; ss; plant; cold tolerance; growth rate; cell cycle pathway;
KW drought tolerance; plant disease resistance; galactomannan; lignin;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW homologous recombination; extreme osmotic condition tolerance;
KW pathogen resistance; pest resistance; yield; photosynthesis; seed oil;
KW stress resistance.
XX
OS Gossypium hirsutum.
XX
PN US2004181830-A1.
XX
PD 16-SEP-2004.
XX
PF 29-JAN-2004; 2004US-00767795.
XX
PR 07-MAY-2001; 2001US-00849529.
PR 12-DEC-2001; 2001US-00021323.
XX
XX (KOVA/) KOVALIC D K.
PA (ZHOU/) ZHOU Y.
PA (CAO/) CAO Y.
PI Kovalic DK, Zhou Y, Cao Y;
XX
DR WPI; 2004-667718/65.
XX
PT New recombinant nucleic acid molecules and polypeptides from Gossypium
PT hirsutum, useful for producing plants with improved biological
PT characteristics (e.g. improved plant cold or drought tolerance).
XX
PS Claim 1; SEQ ID NO 6234; 14pp; English.
XX
CC The invention relates to a recombinant polynucleotide comprising any of
CC the 58798 Cotton plant cDNA sequences mentioned in the specification.
CC Also a recombinant polypeptide comprising any of the 58798 amino acid
CC sequences mentioned in the specification and producing a plant having an
CC improved property. Producing a plant having an improved property
CC comprises transforming a plant with a recombinant construct comprising a
CC promoter region functional in a plant cell operably joined to a
CC polynucleotide comprising a coding sequence for a polypeptide associated
CC with the property, and growing the transformed plant. The polypeptide is
CC useful for improving plant cold tolerance, manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, improving plant
CC drought tolerance, providing increased resistance to plant disease,
CC producing galactomannan (or lignin or plant growth regulators), improving
CC plant heat tolerance, improving plant tolerance to herbicides, increasing
CC the rate of homologous recombination in plants, improving plant tolerance
CC to extreme osmotic conditions or to pathogens or pests, improving yield
CC by modification of photosynthesis, modifying seed oil or protein yield
CC and/or content, improving yield by modification of carbohydrate, nitrogen
CC or phosphorus use and/or uptake, or improving yield by providing improved
CC plant growth and development under at least one stress condition. The
CC polynucleotide and polypeptide may also be used in recombinant DNA
CC constructs, in physical arrays of molecules, as plant breeding markers,
CC or in computer-based storage and analysis systems. The present sequence
CC is a Cotton plant cDNA of the invention. NOTE: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040181830. However only 6585
CC polynucleotide sequences were available, the remaining 5213
CC polynucleotides and all 58798 protein sequences were not present.
XX
SQ Sequence 818 BP; 193 A; 219 C; 164 G; 242 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 13; Length 818;
Best Local Similarity 86.4%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCGCCCTTGAAT 22
DB 192 ATCCTTGCCCGCCCTTGAAT 213
RESULT 14
ABN74461/c
ID ABN74461 standard; cDNA; 1051 BP.
XX
AC ABN74461;
XX
DT 03-JUL-2002 (first entry)
XX
DE Bovine embryonic germ (EG) cell cDNA EST #512.
XX
KW Bovine; Bos taurus; EST; expressed sequence tag; totipotency;
KW development; gene; ss.
XX
OS Bos taurus.
XX
PN WO200194550-A2.
XX
PD 13-DEC-2001.
XX
PF 07-JUN-2001; 2001WO-US018576.
PR 07-JUN-2000; 2000US-0209874P.
PR 06-JUN-2001; 2001US-00876143.
XX
PA (INFI-) INFIGN INC.
XX
PI Eilertsen KJ, Pfister-Genskow M, Childs L;
XX
DR WPI; 2002-351289/38.
XX
PT An expressed sequence tag (EST), the expression of which, or its
PT complementary sequence, in a cell identifies the cell as a
PT developmentally competent or incompetent cell.
XX
PS Example 16; Page 441-442; 584pp; English.
XX
CC The present invention describes an expressed sequence tag (EST), where
CC the EST is an isolated, enriched, or purified nucleic acid sequence
CC representing all or part of a gene, the expression of which, or its
CC complementary sequence, in a cell identifies the cell as a
CC developmentally competent or incompetent cell. Molecules which induce
CC developmental competence in a cell line are useful for inducing
CC totipotency in one or more cells. Molecules which induce developmental
CC incompetence in a cell line are useful for preventing a full term
CC pregnancy in an animal and inhibiting totipotency. The molecules are also
CC useful for treating a disease in an animal by inducing development of one
CC or more cells of the animal into a specific cell type. The present
CC sequence represents a bovine EST which is given in the exemplification of
CC the present invention
SQ Sequence 1051 BP; 208 A; 279 C; 242 G; 236 T; 0 U; 86 Other;

Query Match 77.3%; Score 17; DB 6; Length 1051;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAAT 21
DB 752 TCCTTGCCCGCCCTTGAAT 733
RESULT 15
ABN20125/c

ID	ABN20125 standard; cDNA; 347 BP.	ABN20125 standard; cDNA; 347 BP.	ABN20125 standard; cDNA; 347 BP.
AC	ABN20125;	ABN20125;	ABN20125;
XX			
DT	24-JUN-2002 (first entry)	24-JUN-2002 (first entry)	24-JUN-2002 (first entry)
XX			
DE	Human ORFX polynucleotide sequence SEQ ID NO:8727.	Human ORFX polynucleotide sequence SEQ ID NO:8727.	Human ORFX polynucleotide sequence SEQ ID NO:8727.
XX			
KW	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW	hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;	hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;	hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW	degenerative disorder; osteoarthritis; neurodegenerative disorder;	degenerative disorder; osteoarthritis; neurodegenerative disorder;	degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW	hypertension; hypothyroidism; cholesterol ester storage disease;	hypertension; hypothyroidism; cholesterol ester storage disease;	hypertension; hypothyroidism; cholesterol ester storage disease;
KW	immune deficiency; immune disorder; infectious disease;	immune deficiency; immune disorder; infectious disease;	immune deficiency; immune disorder; infectious disease;
KW	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW	myasthenia gravis; gene; ss.	myasthenia gravis; gene; ss.	myasthenia gravis; gene; ss.
OS	Homo sapiens.	Homo sapiens.	Homo sapiens.
XX			
PN	WO200192523-A2.	WO200192523-A2.	WO200192523-A2.
XX			
PD	06-DEC-2001.	06-DEC-2001.	06-DEC-2001.
XX			
PF	29-MAY-2001; 2001WO-US010836.	29-MAY-2001; 2001WO-US010836.	29-MAY-2001; 2001WO-US010836.
XX			
PR	30-MAY-2000; 2000US-0206132P.	30-MAY-2000; 2000US-0206132P.	30-MAY-2000; 2000US-0206132P.
PR	29-AUG-2000; 2000US-0228716P.	29-AUG-2000; 2000US-0228716P.	29-AUG-2000; 2000US-0228716P.
XX			
PA	(CURA-) CURAGEN CORP.	(CURA-) CURAGEN CORP.	(CURA-) CURAGEN CORP.
XX			
PI	Shinketsu RA, Leach MD;	Shinketsu RA, Leach MD;	Shinketsu RA, Leach MD;
XX			
DR	WPI; 2002-106308/14.	WPI; 2002-106308/14.	WPI; 2002-106308/14.
DR	P-PSDB; ABP04373.	P-PSDB; ABP04373.	P-PSDB; ABP04373.
XX			
PT	Novel human polypeptides and polynucleotides useful for diagnosing,	Novel human polypeptides and polynucleotides useful for diagnosing,	Novel human polypeptides and polynucleotides useful for diagnosing,
PT	preventing and treating cardiovascular disease, neurodegenerative,	preventing and treating cardiovascular disease, neurodegenerative,	preventing and treating cardiovascular disease, neurodegenerative,
PT	hyperproliferative disorders and autoimmune disorders.	hyperproliferative disorders and autoimmune disorders.	hyperproliferative disorders and autoimmune disorders.
XX			
PS	Disclosure; SEQ ID NO 8727; 1037pp; English.	Disclosure; SEQ ID NO 8727; 1037pp; English.	Disclosure; SEQ ID NO 8727; 1037pp; English.
XX			
CC	The present invention describes substantially purified human proteins	The present invention describes substantially purified human proteins	The present invention describes substantially purified human proteins
CC	(referred to as open reading frame, ORFX, where X is 1-11491 (see table 1	(referred to as open reading frame, ORFX, where X is 1-11491 (see table 1	(referred to as open reading frame, ORFX, where X is 1-11491 (see table 1
CC	in the specification). ABN15762 to ABN27252 encode the human ORFX	in the specification). ABN15762 to ABN27252 encode the human ORFX	in the specification). ABN15762 to ABN27252 encode the human ORFX
CC	proteins given in ABP0010 to ABP11500. ORFX proteins are useful for	proteins given in ABP0010 to ABP11500. ORFX proteins are useful for	proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
CC	treating or preventing a pathology associated with an ORFX-associated	treating or preventing a pathology associated with an ORFX-associated	treating or preventing a pathology associated with an ORFX-associated
CC	disorder in humans, and in the manufacture of a medicament for treating a	disorder in humans, and in the manufacture of a medicament for treating a	disorder in humans, and in the manufacture of a medicament for treating a
CC	syndrome associated with ORFX-associated disorder. ORFX polynucleotide	syndrome associated with ORFX-associated disorder. ORFX polynucleotide	syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC	sequences can be used in gene therapy. ORFX sequences can be used in the	sequences can be used in gene therapy. ORFX sequences can be used in the	sequences can be used in gene therapy. ORFX sequences can be used in the
CC	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC	osteoarthritis, neurodegenerative disorders, disorders related to organ	osteoarthritis, neurodegenerative disorders, disorders related to organ	osteoarthritis, neurodegenerative disorders, disorders related to organ
CC	transplantation, cardiovascular diseases, diabetes mellitus, systemic	transplantation, cardiovascular diseases, diabetes mellitus, systemic	transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC	storage disease, various immune deficiencies and disorders, infectious	storage disease, various immune deficiencies and disorders, infectious	storage disease, various immune deficiencies and disorders, infectious
CC	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC	disease and autoimmune inflammatory eye disease. ORFX proteins are also	disease and autoimmune inflammatory eye disease. ORFX proteins are also	disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC	useful for treating burns, incisions, ulcers, for treating osteoporosis,	useful for treating burns, incisions, ulcers, for treating osteoporosis,	useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC	bone degenerative disorders, or periodontal disease, and for gut	bone degenerative disorders, or periodontal disease, and for gut	bone degenerative disorders, or periodontal disease, and for gut
CC	protection or regeneration and treatment of lung or liver fibrosis,	protection or regeneration and treatment of lung or liver fibrosis,	protection or regeneration and treatment of lung or liver fibrosis,
CC	reperfusion injury in various tissues and conditions resulting from	reperfusion injury in various tissues and conditions resulting from	reperfusion injury in various tissues and conditions resulting from
CC	systemic cytokine damage. N.B. The sequence data for this patent did not	systemic cytokine damage. N.B. The sequence data for this patent did not	systemic cytokine damage. N.B. The sequence data for this patent did not
CC	form part of the printed specification, but was obtained in electronic	form part of the printed specification, but was obtained in electronic	form part of the printed specification, but was obtained in electronic
CC	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX			
SQ	Sequence 347 BP; 92 A; 72 C; 99 G; 84 T; 0 U; 0 Other;	Sequence 347 BP; 92 A; 72 C; 99 G; 84 T; 0 U; 0 Other;	Sequence 347 BP; 92 A; 72 C; 99 G; 84 T; 0 U; 0 Other;
	Query Match	Query Match	Query Match
	Best Local Similarity 90.0%; Score 16.8; DB 6; Length 347;	Best Local Similarity 90.0%; Score 16.8; DB 6; Length 347;	Best Local Similarity 90.0%; Score 16.8; DB 6; Length 347;
	Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	2 TCCTTGCCCGCCCTTGAA 21	2 TCCTTGCCCGCCCTTGAA 21	2 TCCTTGCCCGCCCTTGAA 21

Db	84 TCCTTGCCCGCCCTTGGA 65	84 TCCTTGCCCGCCCTTGGA 65
RESULT 16		
AA07266		
ID	AA07266 standard; DNA; 596 BP.	AA07266 standard; DNA; 596 BP.
XX		
AC	AA07266;	AA07266;
XX		
DT	21-NOV-2001 (first entry)	21-NOV-2001 (first entry)
XX		
DE	Human reproductive system related antigen DNA SEQ ID NO: 9954.	Human reproductive system related antigen DNA SEQ ID NO: 9954.
XX		
KW	Human; reproductive system related antigen; reproductive system disorder;	Human; reproductive system related antigen; reproductive system disorder;
KW	cancer; gene therapy; ds.	cancer; gene therapy; ds.
OS	Homo sapiens.	Homo sapiens.
XX		
PN	WO200155320-A2.	WO200155320-A2.
XX		
PD	02-AUG-2001.	02-AUG-2001.
XX		
PF	17-JAN-2001; 2001WO-US001339.	17-JAN-2001; 2001WO-US001339.
XX		
PR	31-JAN-2000; 2000US-0179065P.	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217496P.	11-JUL-2000; 2000US-0217496P.
PR	14-JUL-2000; 2000US-0218290P.	14-JUL-2000; 2000US-0218290P.
PR	26-JUL-2000; 2000US-0220963P.	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225447P.	14-AUG-2000; 2000US-0225447P.
PR	14-AUG-2000; 2000US-0225757P.	14-AUG-2000; 2000US-0225757P.
PR	14-AUG-2000; 2000US-0225758P.	14-AUG-2000; 2000US-0225758P.
PR	14-AUG-2000; 2000US-0225759P.	14-AUG-2000; 2000US-0225759P.
PR	18-AUG-2000; 2000US-0226279P.	18-AUG-2000; 2000US-0226279P.
PR	22-AUG-2000; 2000US-0226681P.	22-AUG-2000; 2000US-0226681P.
PR	22-AUG-2000; 2000US-0226868P.	22-AUG-2000; 2000US-0226868P.
PR	22-AUG-2000; 2000US-0227182P.	22-AUG-2000; 2000US-0227182P.
PR	23-AUG-2000; 2000US-0227009P.	23-AUG-2000; 2000US-0227009P.
PR	30-AUG-2000; 2000US-0228924P.	30-AUG-2000; 2000US-0228924P.
PR	01-SEP-2000; 2000US-0229287P.	01-SEP-2000; 2000US-0229287P.
PR	01-SEP-2000; 2000US-0229343P.	01-SEP-2000; 2000US-0229343P.
PR	01-SEP-2000; 2000US-0229344P.	01-SEP-2000; 2000US-0229344P.
PR	01-SEP-2000; 2000US-0229345P.	01-SEP-2000; 2000US-0229345P.
PR	05-SEP-2000; 2000US-0229509P.	05-SEP-2000; 2000US-0229509P.
PR	05-SEP-2000; 2000US-0229513P.	05-SEP-2000; 2000US-0229513P.
PR	06-SEP-2000; 2000US-0230437P.	06-SEP-2000; 2000US-0230437P.
PR	06-SEP-2000; 2000US-0230438P.	06-SEP-2000; 2000US-0230438P.
PR	08-SEP-2000; 2000US-0231242P.	08-SEP-2000; 2000US-0231242P.
PR	08-SEP-2000; 2000US-0231243P.	08-SEP-2000; 2000US-0231243P.
PR	08-SEP-2000; 2000US-0231244P.	08-SEP-2000; 2000US-0231244P.
PR	08-SEP-2000; 2000US-0231413P.	08-SEP-2000; 2000US-0231413P.

PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225367P.
PR 14-AUG-2000; 2000US-0225368P.
PR 14-AUG-2000; 2000US-0225370P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241828P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483232/52.

Nucleic acids encoding 973 human testicular antigen polypeptides, useful for preventing, diagnosing and/or treating testicular cancer.

Disclosure; SEQ ID NO 3464; 766pp; English.

The present invention provides the protein and coding sequences of 973 human testicular antigens, and fragments of their genomic sequences. The sequences can be used in the treatment of cardiovascular, urinary system, reproductive system, immune, respiratory, neurological and

CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention

SQ Sequence 596 BP; 138 A; 155 C; 178 G; 125 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 4; Length 596;
Best Local Similarity 90.0%; Pred. No. 6.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TCCTTGCCCGCCCTTGAA 21
||||||| |||||
Db 111 TCCTTGCCCGCCCTTGAA 130

RESULT 18

AAAL07268
ID AAL07268 standard; DNA; 618 BP.

XX
AC AAL07268;

DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 9956.

DE Human reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215133P.

PR 07-JUL-2000; 2000US-0216647P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.

```
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 956; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 618 BP; 141 A; 162 C; 186 G; 129 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 4; Length 618;
Best Local Similarity 90.0%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 TCCTTGCCCGCCCTTGAA 21
DB 133 TCCTTGCCCTGCGCTTGAA 152
|||||||
RESULT 19
ABL98814
ID ABL98814 standard; DNA; 618 BP.
XX
AC ABL98814;
XX
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 3466.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200155317-A2.
```

```
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001329.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
```

29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI
XX

DR WPI; 2001-483232/52.
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT for preventing, diagnosing and/or treating testicular cancer.
XX
PS Disclosure; SEQ ID NO 3466; 766pp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention
XX
SQ Sequence 618 BP; 141 A; 162 C; 186 G; 129 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 4; Length 618;
Best Local Similarity 90.0%; Pred. NO. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCGCTTGAA 21
DB 133 TCCTTGCCCGCCGCTTGAA 152

RESULT 20
AAL07267
ID AAL07267 standard; DNA; 621 BP.
XX
AC AAL07267;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9955.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214896P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.

```
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232402P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234984P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Baraash SC, Ruben SM;
XX WPI; 2001-465570/50.
DR
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 9955; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a genomic sequence encoding a
protein of the invention
XX
SQ Sequence 621 BP; 137 A; 164 C; 190 G; 130 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 4; Length 621;
Best Local Similarity 90.0%; Pred No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAA 21
Db 136 TCCTTGCCCGCCCTTGAA 155
RESULT 21
ABL98813
ID ABL98813 standard; DNA; 621 BP.
XX
AC ABL98813;
XX
DT 21-JUN-2002 (first entry)
```

XX Human testicular antigen encoding DNA fragment SEQ ID NO: 3465.
DE Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ds.
XX Homo sapiens.
XX WO200155317-A2.
PD 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001329.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246534P.
PR 08-NOV-2000; 2000US-0246535P.
PR 08-NOV-2000; 2000US-0246536P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.

PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483232/52.
 XX
 XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
 PT for preventing, diagnosing and/or treating testicular cancer.
 XX
 XX Disclosure; SEQ ID NO 3465; 766pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer.
 CC especially testicular cancers. The present sequence is a DNA encoding a
 CC protein fragment of the invention
 XX
 XX Sequence 621 BP; 137 A; 164 C; 190 G; 130 T; 0 U; 0 Other;
 SQ
 Query Match 76.4%; Score 16.8; DB 4; Length 621;
 Best Local Similarity 90.0%; Pred. No. 6.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TCCTTGCCCGCCCTTGAA 21
 Db 136 TCCTTGCCCTGCGCCTTGAA 155
 RESULT 22
 ABI99929/c
 ID ABI99929 standard; cDNA; 716 BP.
 XX
 AC ABI99929;
 XX
 DT 07-AUG-2003 (revised)
 DT 08-MAR-2002 (first entry)
 XX
 DE Rat mucocardial cell proliferation associated cDNA SEQ ID NO 29.
 XX
 XX Rat; heart; cardiant; myocardial necrosis; cardiac hypertrophy;
 XX cardiac insufficiency; ss.
 OS Rattus norvegicus.
 XX
 PN WO200183705-A1.
 XX
 PD 08-NOV-2001.
 XX
 PF 07-AUG-2003 (revised)
 DT 08-MAR-2002 (first entry)
 XX
 DE Rat mucocardial cell proliferation associated cDNA SEQ ID NO 29.
 XX
 XX Rat; heart; cardiant; myocardial necrosis; cardiac hypertrophy;
 XX cardiac insufficiency; ss.
 OS Rattus norvegicus.
 XX
 PN WO200183705-A1.
 XX
 PD 08-NOV-2001.
 XX
 PF 27-APR-2001; 2001WO-JP003700.
 XX
 PR 27-APR-2000; 2000JP-00126741.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Yamada Y, Sekine S, Kikuchi Y, Sakurada K;
 XX WPI; 2002-075160/10.
 XX
 XX Genes having differential expression in fetal and adult heart tissue
 PT useful for screening potential drugs for promoting repair of damage
 PT caused by myocardial necrosis.
 XX
 XX Example 4; Page 143-144; 171pp; Japanese.

CC The invention relates to gene sequences (ABI99915-ABI99934) having
 CC modified expression in fetal heart tissue as compared to adult heart
 CC tissue and the encoded proteins (ABB57375-ABB57392). The genes have
 CC cardiant activity and may be useful in the promotion of the repair of
 CC damage to heart tissue caused by myocardial necrosis. The gene sequences
 CC are useful for screening potential compounds for the ability to influence
 CC disease associated with myocardial necrosis. Drugs identified by the
 CC screening methods may be used to treat and prevent disease with which
 CC myocardial necrosis is associated, such as cardiac hypertrophy and
 CC cardiac insufficiency. Diagnosis of diseases such as those above is also
 CC disclosed. (Updated on 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 716 BP; 178 A; 189 C; 213 G; 136 T; 0 U; 0 Other;
 Query Match 76.4%; Score 16.8; DB 6; Length 716;
 Best Local Similarity 90.0%; Pred. No. 6.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 CCTTGCCCGCCCTTGAA 22
 Db 166 CCTTGCCCGCCCTTGAGT 147
 RESULT 23
 ABI99930/c
 ID ABI99930 standard; cDNA; 787 BP.
 XX
 AC ABI99930;
 XX
 DT 07-AUG-2003 (revised)
 DT 08-MAR-2002 (first entry)
 XX
 DE Rat mucocardial cell proliferation associated cDNA SEQ ID NO 30.
 XX
 XX Rat; heart; cardiant; myocardial necrosis; cardiac hypertrophy;
 XX cardiac insufficiency; ss.
 OS Rattus norvegicus.
 XX
 PN WO200183705-A1.
 XX
 PD 08-NOV-2001.
 XX
 PF 27-APR-2001; 2001WO-JP003700.
 XX
 PR 27-APR-2000; 2000JP-00126741.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Yamada Y, Sekine S, Kikuchi Y, Sakurada K;
 XX WPI; 2002-075160/10.
 XX
 XX Genes having differential expression in fetal and adult heart tissue
 PT useful for screening potential drugs for promoting repair of damage
 PT caused by myocardial necrosis.
 XX
 XX Claim 1; Page 144-145; 171pp; Japanese.
 CC
 CC The invention relates to gene sequences (ABI99915-ABI99934) having
 CC modified expression in fetal heart tissue as compared to adult heart
 CC tissue and the encoded proteins (ABB57375-ABB57392). The genes have
 CC cardiant activity and may be useful in the promotion of the repair of
 CC damage to heart tissue caused by myocardial necrosis. The gene sequences
 CC are useful for screening potential compounds for the ability to influence
 CC disease associated with myocardial necrosis. Drugs identified by the
 CC screening methods may be used to treat and prevent disease with which
 CC myocardial necrosis is associated, such as cardiac hypertrophy and
 CC cardiac insufficiency. Diagnosis of diseases such as those above is also
 CC disclosed. (Updated on 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 787 BP; 193 A; 208 C; 236 G; 150 T; 0 U; 0 Other;

```
Query Match          76.4%; Score 16.8; DB 6; Length 787;
Best Local Similarity 90.0%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTGCCCCGCCCCCTTGAAT 22
Db 237 CCTGCCCCGCCCCCTTGAAT 218

RESULT 24
ABK49702/c
ID ABK49702 standard; cDNA; 1126 BP.
XX
AC ABK49702;
XX
DT 15-JUL-2002 (first entry)
XX
DE cDNA encoding human kinase uridylylate 12 protein.
XX
KW Human; kinase uridylylate 12; malignant neoplasm; tumour; haemopathy; HIV;
KW human immunodeficiency virus infection; immunological disease;
KW inflammation; sterility; pregnancy pathology; growth disorder;
KW embryonic development disorder; hereditary disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 377..712
FT /tag= a
FT /product= "Human kinase uridylylate 12 protein"
XX
PN WO200183726-A1.
XX
PD 08-NOV-2001.
XX
PF 28-APR-2001; 2001WO-CN000649.
XX
PR 29-APR-2000; 2000CN-00115551.
XX
PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI Mao Y, Xie Y;
XX
WPI: 2002-062122/08.
DR P-PSDB; AAU79947.
XX
Human uridylylate kinase 12 and encoded polynucleotide, applicable in
PT diagnosis and treatment of developmental disorders, malignant tumor,
PT hemopathy, HIV infection, immunological diseases and various
PT inflammations.
XX
PS Claim 6; Page 29-30; 36pp; Chinese.
XX
The present invention relates to a new polypeptide of human uridylylate
CC kinase 12, its fragment, analogue or derivative. The polypeptide of the
CC invention and its encoding polynucleotide are applicable in diagnosis and
CC treatment of malignant neoplasm, haemopathy, human immunodeficiency virus
CC (HIV) infection, immunological diseases, various inflammations,
CC sterility, pregnancy pathology, embryonic development disorders, growth
CC and development disorders and hereditary diseases. The present nucleic
CC acid sequence encodes the human kinase uridylylate 12 protein of the
CC invention
XX
SQ Sequence 1126 BP; 313 A; 277 C; 244 G; 292 T; 0 U; 0 Other;

Query Match          76.4%; Score 16.8; DB 6; Length 1126;
Best Local Similarity 90.0%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTGCCCCGCCCCCTTGAAT 22
Db 267 CCTGCTCGGCCCCCTTGAAT 248

RESULT 25
AAK94185
ID AAK94185 standard; cDNA; 2030 BP.
XX
AC AAK94185;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA; cDNA synthesis; oligo-capping; ss.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
DR P-PSDB; AAM93269.
XX
830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
Claim 8; SEQ ID NO 2732; 1380pp + Sequence Listing; English.
XX
The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2030 BP; 479 A; 514 C; 635 G; 402 T; 0 U; 0 Other;

Query Match          76.4%; Score 16.8; DB 4; Length 2030;
Best Local Similarity 90.0%; Pred. No. 7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCCCTTGAAT 21
Db 1544 TCCTTGCCCCGCCCCCTTGAAT 1563

RESULT 26
ADL30699
ID ADL30699 standard; cDNA; 2030 BP.
XX
AC ADL30699;
XX
DT 20-MAY-2004 (first entry)
XX
DE Full length human cDNA clone SeqID 2732.
XX
```

KW human; medicine; signal transduction; glycoprotein; transcription;
 KW oligo-capping method; ss; gene.
 OS Homo sapiens.
 XX EP1396543-A2.
 PN 10-MAR-2004.
 XX 07-JUL-2000; 2003BP-00025638.
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00187774.
 PR 02-MAY-2000; 2000JP-00183865.
 XX 07-JUL-2000; 2000EP-00114089.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 PA Ota T, Nishikawa T, Isegai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2004-204755/20.
 DR P-PSDB; ADL30700.
 XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 PT length human cDNAs.
 XX Example 1; SEQ ID NO 2732; 1340pp; English.
 XX This invention relates to a novel primers useful for synthesising full
 CC length cDNA molecules that encode human proteins. Specifically, it refers
 CC to secretory or membrane proteins that are potential therapeutic agents/
 CC target molecules in the field of medicine, and in particular genes
 CC encoding proteins that are associated with signal transduction,
 CC glycoproteins and transcription. The present invention describes a method
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'
 CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.
 XX Sequence 2030 BP; 479 A; 514 C; 635 G; 402 T; 0 U; 0 Other;
 SQ Query Match 76.4%; Score 16.8; DB 12; Length 2030;
 Best Local Similarity 90.0%; Pred. No. 7e+02; Mismatches 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 TCCTTGCCCGCCCTTGAA 21
 DB 1544 TCCTTGCCCTGCGCCCTTGAA 1563
 RESULT 27
 ADQ68003
 ID ADQ68003 standard; cDNA; 2030 BP.
 XX ADQ68003;
 AC ADQ68003;
 XX 21-OCT-2004 (first entry)
 XX Recombinant vector preparation human clone cDNA, HEMBB1001048.
 DE Recombinant vector; homology search program; storage unit; colony picker;
 XX automatic divider; gene cluster; human; clone; ss.
 KW Homo sapiens.
 OS JP2004215665-A.
 PN 05-AUG-2004.
 XX 26-DEC-2003; 2003JP-00435752.
 PF 27-DEC-2002; 2002JP-00381924.
 XX

PA (ZOLJ-) ZOLJIN KK.
 XX WPI; 2004-557634/54.
 XX Simultaneously preparing several recombinant vectors, by inserting known
 PT DNA fragment of different lengths or types to vectors of same or
 PT different species, identifying recombinant vector having DNA fragment and
 PT aliquoting vector.
 XX Example 1; SEQ ID NO 2; 26pp; Japanese.
 PS The invention relates to a novel method for simultaneously preparing
 CC several recombinant vectors. The method involves inserting an objective
 CC known DNA fragment of different lengths or of different types to vectors
 CC of same species or different types, identifying the recombinant vector in
 CC which the DNA fragment has been inserted present in the clone of a host
 CC cell, which is transformed with the recombinant vector, and aliquoting
 CC the recombinant vector. The invention further comprises: a method for
 CC identifying recombinant vectors in which DNA fragment is inserted; a
 CC computer program for performing the procedure of determining the
 CC similarity between the sequence data using a homology search program, and a
 CC input unit for inputting the sequence data of a recombinant vector, and a
 CC storage unit which inputs and stores the target DNA sequence and type of
 CC target DNA; displaying the result of the above method, where the results
 CC are displayed visually, corresponding to the format of the container
 CC which performs the DNA sequencing of the grown colony; and an output
 CC method for the result from the above method, where the result is output
 CC in a colony picker or format corresponding to an automatic divider. The
 CC method is an efficient method for the insertion of a gene cluster into a
 CC vector. This polynucleotide represents a human clone cDNA sequence used
 CC in the recombinant vector preparation method of the invention.
 XX Sequence 2030 BP; 479 A; 514 C; 635 G; 402 T; 0 U; 0 Other;
 SQ Query Match 76.4%; Score 16.8; DB 13; Length 2030;
 Best Local Similarity 90.0%; Pred. No. 7e+02; Mismatches 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 TCCTTGCCCGCCCTTGAA 21
 DB 1544 TCCTTGCCCTGCGCCCTTGAA 1563
 RESULT 28
 ABL23652/c
 ID ABL23652 standard; DNA; 2513 BP.
 XX ABL23652;
 AC ABL23652;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 22429.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
PS Claim 1; SEQ ID NO 22429; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2513 BP; 693 A; 532 C; 637 G; 651 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 4; Length 2513;
Best Local Similarity 90.0%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAA 21
||||| ||||| |||||
Db 331 TCCTTGGCCCGCCCTTGAA 312

RESULT 29
ABL20443/c
ID ABL20443 standard; DNA; 4013 BP.
XX
AC ABL20443;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12802.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 12802; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4013 BP; 1096 A; 1067 C; 1062 G; 788 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 4; Length 4013;
Best Local Similarity 90.0%; Pred. No. 7.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20
||||| ||||| |||||
Db 667 ATCCTTGCCCGCCCTTGA 648

RESULT 30
ABL18683/c
ID ABL18683 standard; DNA; 4098 BP.
XX
AC ABL18683;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7522.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 7522; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4098 BP; 1123 A; 1081 C; 1075 G; 819 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 4; Length 4098;
Best Local Similarity 90.0%; Pred. No. 7.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20
||||| ||||| |||||
Db 752 ATCCTTGCCCGCCCTTGA 733

RESULT 31
ABL18682

XX WPI; 1998-296771/26.
 XX Mycobacterial shuttle plasmids comprising mycobacteriophage L5 with
 PT lambda cosmid insert - useful for delivery of foreign DNA into
 PT mycobacteria and for generating mutations.
 XX
 PS Claim 7; Col 5-28; 42pp; English.
 XX
 CC The present sequence represents a L5 shuttle plasmid designated pHA41.
 CC It was prepared by inserting an Escherichia coli bacteriophage cosmid,
 CC pYUB328, into a non-essential region of an L5 mycobacteriophage genome.
 CC Foreign DNA such as reporter genes (e.g. luciferase), transposons (e.g.
 CC IS1096) and mycobacterial inhibitor genes can also be inserted into the
 CC phasmid. The phasmids can be used to introduce foreign DNA into
 CC mycobacteria. The L5 shuttle plasmid can be used to introduce mutations
 CC that can be used to study the mechanisms of mycobacteria and to develop
 CC vaccines and drugs for treating mycobacterial infections. (Updated on 27-
 CC AUG-2003 to correct OS field.)
 XX
 SQ Sequence 50341 BP; 9641 A; 15673 C; 15276 G; 9750 T; 0 U; 1 Other;
 Query Match 76.4%; Score 16.8; DB 2; Length 50341;
 Best Local Similarity 90.0%; Pred. No. 8.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATCCTTGCCCGCCCTTGA 20
 |||||
 Db 7079 ATCCTTGCCCGCCCTTGA 7060
 |||||
 RESULT 34
 AAZ39519/c
 ID AAZ39519 standard; DNA; 50341 BP.
 XX
 AC AAZ39519;
 XX
 DT 06-AUG-2003 (revised)
 DT 11-FEB-2000 (first entry)
 XX
 DE L5 shuttle plasmid pHA41 sequence.
 XX
 KW Mycobacterium; mutation; L5 shuttle plasmid; mycobacteriophage L5;
 KW transposon; tuberculosis; gene analysis; shuttle mutagenesis; ss.
 XX
 OS Mycobacterium phase L5.
 XX
 PN US5994137-A.
 XX
 PD 30-NOV-1999.
 XX
 PF 11-MAY-1998; 98US-00075904.
 XX
 PR 07-FEB-1992; 92US-00833431.
 PR 29-APR-1993; 93US-00057531.
 PR 23-MAY-1994; 94US-00247901.
 XX
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 PA (UYPI-) UNIV PITTSBURGH.
 XX
 PI Mcadam R, Bardarov S, Jacobs WR, Hatfull GF;
 XX
 DR WPI; 2000-038266/03.
 XX
 XX Generation of mycobacterial mutations with a phasmid, useful for
 PT producing mycobacterial vaccines.
 XX
 PS Example 2; Col 5-48; 71pp; English.
 XX
 CC The invention is directed towards generating mycobacterial mutations
 CC using an L5 shuttle plasmid. The method comprises: (a) obtaining an L5
 CC shuttle plasmid comprising mycobacteriophage L5 with an Escherichia coli
 CC bacteriophage lambda cosmid and a transposon inserted into a non-

CC essential region between the PvuII restriction site at nucleotide 42986
 CC and the SacII restriction site at nucleotide 52191 of the
 CC mycobacteriophage L5 genome; and (b) infecting a mycobacterium with the
 CC L5 shuttle plasmid to effect delivery of the transposon from the phasmid
 CC to the chromosome of the mycobacterium, causing a mutation in the gene of
 CC the mycobacterium. The creation of mutants in Mycobacterium tuberculosis
 CC is essential in the analysis of genes, and can be used to produce
 CC mycobacterial vaccines for treating tuberculosis. The method is
 CC advantageous compared to the prior art of shuttle mutagenesis since it
 CC does not require homologous recombination. The present sequence
 CC represents the L5 shuttle phasmid pHA41 sequence. (Updated on 06-AUG-
 CC 2003 to correct OS field.)
 XX
 SQ Sequence 50341 BP; 9641 A; 15673 C; 15276 G; 9750 T; 0 U; 1 Other;
 Query Match 76.4%; Score 16.8; DB 3; Length 50341;
 Best Local Similarity 90.0%; Pred. No. 8.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATCCTTGCCCGCCCTTGA 20
 |||||
 Db 7079 ATCCTTGCCCGCCCTTGA 7060
 |||||
 RESULT 35
 AAT51411/c
 ID AAT51411 standard; DNA; 52297 BP.
 XX
 AC AAT51411;
 XX
 DT 27-AUG-2003 (revised)
 DT 23-APR-1997 (first entry)
 XX
 DE Mycobacteriophage L5 genome sequence.
 XX
 KW Reporter mycobacteriophage; Mycobacterium tuberculosis; infection;
 KW drug susceptibility; diagnosis; ss.
 XX
 OS Mycobacterium phase L5.
 XX
 PN WO9425572-A1.
 XX
 PD 10-NOV-1994.
 XX
 PF 29-APR-1994; 94WO-US004788.
 XX
 PR 29-APR-1993; 93US-00057531.
 XX
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 PA (UYPI-) UNIV PITTSBURGH.
 XX
 PI Jacobs WR, Bloom BR, Hatfull GF;
 XX
 DR WPI; 1995-022252/03.
 XX
 PT Reporter myco-bacteriophages - useful for rapid diagnosis of myco-
 PT bacterial infection and assessment of drug susceptibilities to
 PT mycobacterial strains.
 XX
 PS Disclosure; Page 103-137; 192pp; English.
 XX
 CC The entire genome sequence of mycobacteriophage L5 is given in AAT51411.
 CC L5 is a temperate virus with broad host range among mycobacteria.
 CC Transcriptional promoters and reporter genes (esp. luciferase or beta-
 CC galactosidase) may be introduced into the genome sequence to provide
 CC mycobacterial species-specific reporter mycobacteriophages, e.g. pHG51
 CC (ATCC 75454) and pHG55 (ATCC 75453). These are useful in the rapid
 CC diagnosis of mycobacterial infection and for the assessment of drug
 CC susceptibilities of mycobacteria in clinical samples. Specifically, a
 CC reporter mycobacteriophage can be used to diagnose tuberculosis and to
 CC assess the drug susceptibilities of various strains of Mycobacterium
 CC tuberculosis. (Updated on 27-AUG-2003 to correct OS field.)
 XX

SQ Sequence 52297 BP; 9633 A; 16532 C; 16025 G; 10107 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 2; Length 52297;
Best Local Similarity 90.0%; Pred. No. 8.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCGCCCTTGA 20
DB 6976 ATCCTTGCCCGCCCTTGA 6957

RESULT 36
AAQ47357/c
ID AAQ47357 standard; DNA; 52298 BP.
XX
AC AAQ47357;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 07-JUN-1994 (first entry)
XX
DE L5 mycobacteriophage DNA.
XX
KW Genome; L5: mycobacteriophage; temperate virus; mycobacteria; phage;
KW phage g; lysogeny; integration; chromosome; attP; attachment site;
KW 62-61-60 gene; left arm; integrase; integration-proficient vector;
KW reporter gene; diagnosis; right arm; tuberculosis; drug resistance;
KW mycobacterial species-specific reporter mycobacteriophage; ss.
XX
OS Mycobacterium phage L5.
XX
PN WO9316172-A1.
XX
PD 19-AUG-1993.
XX
PF 02-FEB-1993; 93WO-US000913.
XX
PR 07-FEB-1992; 92US-00833431.
XX
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PA (UYPI-) UNIV PITTSBURGH.
XX
PI Jacobs WR, Bloom BR, Hatfull GF;
XX
PS WPI; 1993-272877/34.
XX
PT Mycobacterial species-specific reporter myco-bacteriophages - used to
PT detect infection by or drug resistance of mycobacteria.
XX
PS Disclosure; Page 11-25; 65pp; English.
XX
CC This sequence represents the DNA sequence of the entire genome of the L5
CC mycobacteriophage. L5 is a temperate virus with a broad host range among
CC mycobacteria, and is morphologically similar to the family of phages that
CC includes phage 9 and contain a linear double stranded genome with
CC cohesive ends. During the establishment of lysogeny, the L5 genome
CC becomes integrated into the mycobacterial chromosome at the attP
CC attachment site. DNA analysis has indicated that the L5 genome is
CC organised into a right and left arm with the attachment site and
CC integrase at the center of the genome. The L5 genome may be used for the
CC production of integration-proficient vectors for mycobacteria. Part of
CC the L5 genome is not essential for mycobacteriophage growth. It has been
CC demonstrated that all or part of the gene 62-61-60 can be deleted without
CC affecting the life cycle of the L5 phage. This region is therefore
CC suitable for insertion of reporter genes. Mycobacterial species-specific
CC reporter mycobacterio- phages produced in this way may be used to
CC diagnose mycobacterial disease, preferably tuberculosis, by incubating
CC with a sample and detecting the reporter gene product. They may also be
CC used for assessing drug resistance of a mycobacterial strain by adding
CC the reporter vector to the strain, then adding the drug and detecting the
CC formation of any gene product. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX

SQ Sequence 52298 BP; 9631 A; 16531 C; 16022 G; 10106 T; 0 U; 8 Other;
Query Match 76.4%; Score 16.8; DB 2; Length 52298;
Best Local Similarity 90.0%; Pred. No. 8.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCGCCCTTGA 20
DB 6976 ATCCTTGCCCGCCCTTGA 6957

RESULT 37
ADQ97316/c
ID ADQ97316 standard; DNA; 107330 BP.
XX
AC ADQ97316;
XX
DT 07-OCT-2004 (first entry)
XX
DE Mouse cancer associated sequence MD08-031, SEQ ID 293.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS Mus musculus.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 293; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 107330 BP; 30385 A; 20267 C; 20584 G; 34511 T; 0 U; 1583 Other;
Query Match 76.4%; Score 16.8; DB 12; Length 107330;
Best Local Similarity 90.0%; Pred. No. 9.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGA 21
DB 23142 TCCTTGCCCGCCCTTGA 23123

RESULT 38
ABT05583/c
ID ABT05583 standard; DNA; 330 BP.
XX
AC ABT05583;
XX
DT 24-OCT-2002 (first entry)
XX
DE Mycobacterium tuberculosis promoter sequence 1.
XX

PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 40342; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1256 BP; 395 A; 306 C; 285 G; 270 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 4; Length 1256;
Best Local Similarity 94.4%; Pred. No. 1e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 TCCTTGCCCGCCCTTG 19
|||||
DB 352 TCCTTGCCCTCCCTTG 369
|||||
RESULT 41
ACA35376/c
ID ACA35376 standard; DNA; 1567 BP.
XX ACA35376;
AC ACA35376;
DT 19-JUN-2003 (first entry)
DE Prokaryotic essential gene #17033.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX Klebsiella pneumoniae.
OS WO200277183-A2.
PN 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU31506.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 23246; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification.
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1567 BP; 344 A; 442 C; 420 G; 361 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 8; Length 1567;
Best Local Similarity 94.4%; Pred. No. 1e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 CTGCGCCCGCCCTTGAA 21
|||||
DB 1285 CTGCGCCCGCTCCTTGAA 1268
|||||
RESULT 42
ABL29622/c
ID ABL29622 standard; DNA; 3549 BP.
XX ABL29622;
AC ABL29622;
XX 26-MAR-2002. (first entry)
DT 26-MAR-2002. (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 40339.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
XX

```

PF 23-MAR-2001; 2001WO-US009231.
XX
XX AAV17758/c
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 40339; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3549 BP; 1023 A; 704 C; 774 G; 1048 T; 0 U; 0 Other;
SQ
Query Match 74.5%; Score 16.4; DB 4; Length 3549;
Best Local Similarity 94.4%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
Db 1967 TCCTTGCCCGCCCTTG 1950

RESULT 43
AAV17758/c
ID AAV17758 standard; DNA; 4739 BP.
XX
XX AAV17758;
XX
XX 03-JUL-1998 (first entry)
XX
XX Mycobacterium bovis BCGINV-33 gene.
XX
XX Mycobacterium bovis BCG protein; Mcep; tuberculosis; BCGINV-33;
KW hybridisation test; probe; amplification; ss.
XX
XX Mycobacterium bovis.
XX
XX Key Location/Qualifiers
FH CDS 1802..3383
FT /*tag= a
FT /product= "BCG protein"
FT 2336..3383
FT /*tag= b
FT /product= "Mcep protein"
XX
XX WO9801559-A1.
XX
XX 15-JAN-1998.
XX
XX 09-JUL-1997; 97WO-CA000484.
XX
XX 10-JUL-1996; 96US-00677970.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Anand NN, Klein MH;
PI

```

```

XX WPI; 1998-110232/10.
DR P-PSDB; AAW48359.
XX
XX Nucleic acid encoding mycobacterial protein involved in cell binding and
PT entry - used for diagnosis of Mycobacterium infection and in vaccines for
PT humans or animals.
XX
XX Disclosure; Fig 4A-4C; 107pp; English.
XX
XX The present sequence represents a Mycobacterial bovis BCGINV-33 gene
CC encoding for a protein (45-60 kDa) associated with cell binding and
CC entry. The gene sequence contains two translational frames, one for the
CC M.bovis BCGINV-33 gene and the other for the M.tuberculosis Mce gene
CC (nucleotides 2336-3383). Therefore the Mcep protein is an internal
CC polypeptide of the protein from M.bovis. The invention relates to the use
CC of the BCGINV-33 gene sequence in hybridisation tests for diagnosis of
CC M.tuberculosis infections. Fragments of the BCGINV-33 gene sequence can
CC be used in PCR to detect Mycobacterium in tissues and body fluids and
CC also for isolating related genes. BCG protein (AAW48359) or its fragments
CC can be used in vaccines to generate an immune response that may be
CC capable of protecting humans and animals (especially cattle) against
CC mycobacterial infections
XX
XX Sequence 4739 BP; 851 A; 1440 C; 1503 G; 945 T; 0 U; 0 Other;
SQ
Query Match 74.5%; Score 16.4; DB 2; Length 4739;
Best Local Similarity 94.4%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
Db 282 TCCTTGCCCGCCCTTG 265

RESULT 44
ABL12964
ID ABL12964 standard; cDNA; 32503 BP.
XX
XX ABL12964;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33374.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB68861.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 33374; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC

```

CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 32503 BP; 8879 A; 7155 C; 7979 G; 8490 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 4; Length 32503;
Best Local Similarity 94.4%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATCTTGGCCCGCCCTT 18
|||||
Db 16979 ATCTTGGCCCGACCTT 16996

RESULT 45

ACN44588
ID ACN44588 standard; DNA; 33454 BP.

XX AC ACN44588;

XX DT 18-NOV-2004 (first entry)

XX DE Mouse genomic sequence mcG22056.

XX KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX OS Mus musculus.

XX PN WO2003073826-A2.

XX PD 12-SEP-2003.

XX PF 28-FEB-2003; 2003WO-US006235.

XX PR 01-MAR-2002; 2002US-00087192.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW;

XX DR WPI; 2003-328604/31.

XX PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.

XX PS Claim 1; SEQ ID NO 1111; Opp; English.

XX CC The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published

SQ Sequence 33454 BP; 8053 A; 8504 C; 8362 G; 8515 T; 0 U; 20 Other;

Query Match 74.5%; Score 16.4; DB 11; Length 33454;
Best Local Similarity 94.4%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 CTGTGCCCCCCTTGAA 21
|||||
Db 8425 CATGCCCCCCTTGAA 8442

RESULT 46

AAI99682_01/c
Continuation (2 of 45) of AAI99682 from base 100001 (Mycobacterium tuberculosis strain H;
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 74.5%; Score 16.4; DB 4; Length 110000;
Best Local Similarity 94.4%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCCTTGGCCCGCCCTTG 19
|||||

Db 96793 TCCTTGGCCCGACCTTG 96776

RESULT 47

AAI99683_01/c
Continuation (2 of 44) of AAI99683 from base 100001 (Mycobacterium tuberculosis strain H;
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000

AAI199683_01 100001 210000
AAI199683_02 200001 310000
AAI199683_03 300001 410000
AAI199683_04 400001 510000
AAI199683_05 500001 610000
AAI199683_06 600001 710000
AAI199683_07 700001 810000
AAI199683_08 800001 910000
AAI199683_09 900001 1010000
AAI199683_10 1000001 1110000
AAI199683_11 1100001 1210000
AAI199683_12 1200001 1310000
AAI199683_13 1300001 1410000
AAI199683_14 1400001 1510000
AAI199683_15 1500001 1610000
AAI199683_16 1600001 1710000
AAI199683_17 1700001 1810000
AAI199683_18 1800001 1910000
AAI199683_19 1900001 2010000
AAI199683_20 2000001 2110000
AAI199683_21 2100001 2210000
AAI199683_22 2200001 2310000
AAI199683_23 2300001 2410000
AAI199683_24 2400001 2510000
AAI199683_25 2500001 2610000
AAI199683_26 2600001 2710000
AAI199683_27 2700001 2810000
AAI199683_28 2800001 2910000
AAI199683_29 2900001 3010000
AAI199683_30 3000001 3110000
AAI199683_31 3100001 3210000
AAI199683_32 3200001 3310000
AAI199683_33 3300001 3410000
AAI199683_34 3400001 3510000
AAI199683_35 3500001 3610000
AAI199683_36 3600001 3710000
AAI199683_37 3700001 3810000
AAI199683_38 3800001 3910000
AAI199683_39 3900001 4010000
AAI199683_40 4000001 4110000
AAI199683_41 4100001 4210000
AAI199683_42 4200001 4310000
AAI199683_43 4300001 4403765

Query Match 74.5%; Score 16.4; DB 4; Length 110000;
Best Local Similarity 94.4%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCCTTGCCCGCCCTTG 19
|||||
DB 96960 TCCTTGCCCGCCCTTG 96943

RESULT 48
ADQ97481/c
ID ADQ97481 standard; DNA; 194534 BP.
XX
AC ADQ97481;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD09-002, SEQ ID 458.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX

PR 27-DEC-2002; 2002US-00330773.
XX (SAGR-) SAGRES DISCOVERY INC.
PA
XX Morris DW, Malandro MS;
PI
XX WPI; 2004-543781/52.
DR
XX New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 458; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 194534 BP; 56522 A; 39287 C; 37893 G; 59621 T; 0 U; 1211 Other;
Query Match 74.5%; Score 16.4; DB 12; Length 194534;
Best Local Similarity 94.4%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 CTTGCCCGCCCTTGAA 21
|||||
DB 73115 CTTGCCCGCCCTTGAA 73098

RESULT 49
ACH91211
ID ACH91211 standard; DNA; 155 BP.
XX
AC ACH91211;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon probe #24406.
XX
KW Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX
DR WPI; 2004-119264/12.
XX
PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
PS Claim 1; SEQ ID NO 24406; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and

coding at least 8 amino acids of any of the 6888 amino acid sequences
hybridises under high stringency conditions to a nucleic acid probe that
expressed in human cells or tissues. Also included are a spatially-
addressable set of single exon nucleic acid probes for measuring human
gene expression (comprising a plurality of single exon nucleic acid
probes cited above, where each of the plurality of probes is separately
and addressably isolatable or amplifiable from the plurality), a single
exon microarray for measuring human gene expression, a method of
measuring human gene expression, a vector comprising the single exon
probe cited above, an ORF-encoded peptide comprising at least 8
contiguous amino acids of any of the above-mentioned amino acid
sequences (optionally with conservative amino acid substitutions), an
isolated antibody that binds specifically to a peptide cited above,
methods of selling and/or licensing single exon probes or microarrays to
a customer desiring to measure gene expression, a method of providing
human gene expression data by subscription, and a computer-readable
storage medium which contains a database having a plurality of records
(each record including data on the expression of a single exon probe
cited above. The probe, methods and apparatus are useful in gene
expression analysis. The probes may be used as tools for surveying
tissues to detect the presence of expressed messages that contain their
specific exon, or in constructing genome-derived single exon microarrays.
In addition, the probes are used in identifying and characterising
alternative splicing events, in detecting and characterising gross
alterations in the genomic locus that includes their exon, in assessing
smaller genomic alterations, in priming the synthesis of nucleic acids,
or in expressing the ORF-encoded peptide. The present sequence is a human
single exon probe of the invention. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from USPTO at
seqdata.uspto.gov/sequence.html?DocID=20030194704

XX Sequence 155 BP; 24 A; 56 C; 34 G; 41 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 12; Length 155;

Best Local Similarity 85.7%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22

DB 66 TCCTTGCTTCCTTGAAT 86

RESULT 50

ABK80858/c

ID ABK80858 standard; DNA; 222 BP.

XX AC ABK80858;

XX 13-AUG-2002 (first entry)

XX Bacillus clausii genomic sequence tag (GST) #3701.

XX Differential gene expression; genomic sequenced tag; GST;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

OS Bacillus clausii.

PN WO200229113-A2.

PD 11-APR-2002.

XX 05-OCT-2001; 2001WO-US031437.

XX 06-OCT-2000; 2000US-00680598.

PR 27-MAR-2001; 2001US-0279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

PA (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;

PI

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.

XX Claim 11; SEQ ID NO 8149; 200pp; English.

XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 222 BP; 57 A; 49 C; 38 G; 78 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 6; Length 222;

Best Local Similarity 85.7%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22

DB 192 TCCATGCCCGCTCTTGAAT 172

Search completed: April 25, 2005, 11:36:39

Job time : 238.431 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:13:24 ; Search time 1893.14 Seconds
(without alignments)
442.342 Million cell updates/sec

Title: US-10-010-476-14
Perfect score: 22
Sequence: 1 ATCTTGCCCGCCCTTGAAT 22

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gaa1:
9: gb_gaa2:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	86.4	240	8	CC249763 CH261-16E
2	18.8	85.5	684	6	CA366933 642629 NC
3	18.8	85.5	1158	6	CD497293 CDA27-A09
C 4	18.4	83.6	398	1	AI972057 wv31h04.x
C 5	18.4	83.6	408	2	BF511913 UI-H-B14
C 6	18.4	83.6	436	2	BF430395 933 MARC
C 7	18.4	83.6	440	8	BH828773 BACPP28-O
C 8	18.4	83.6	473	2	BF507848 UI-H-B14
C 9	18.4	83.6	478	1	AI380413 tf96c06.x
C 10	18.4	83.6	638	6	CA397297 C89406.y
C 11	18.4	83.6	722	8	AQ939474 NRI-180R
C 12	18.4	83.6	918	2	BE957426 601633569
C 13	18	81.8	1139	1	AL567601 AL567601
C 14	17.8	80.9	269	4	BM208321 C0625F08-
C 15	17.8	80.9	279	2	BB083676 BB083676
C 16	17.8	80.9	331	4	BM222541 K0122G08-
C 17	17.8	80.9	337	5	BY393360 BY393360
C 18	17.8	80.9	338	1	AA959728 wv54c07.8
C 19	17.8	80.9	339	1	AA959831 wv54g07.8
C 20	17.8	80.9	340	1	AI646086 wv54g02.x
C 21	17.8	80.9	348	6	CB704661 AMGNNUC.N
22	17.8	80.9	352	1	AA032751 ml35a02.f
23	17.8	80.9	357	5	BY400112 BY400112
24	17.8	80.9	361	5	BY467320 BY467320

c 98	17.8	80.9	644	5	BQ747303	UI-M-FAO-	BQ747303	UI-M-FAO-	c 171	17.4	79.1	1059	5	BX355685	BX355685
c 99	17.8	80.9	655	7	CO089153	GR_Ea08I	CO089153	GR_Ea08I	c 172	17.4	79.1	1074	6	CD556201	CD556201
c 100	17.8	80.9	657	7	CF902322	AO339C06	CF902322	AO339C06	c 173	17.2	78.2	207	2	BB593516	BB593516
c 101	17.8	80.9	659	7	CO087059	GR_Ea05F	CO087059	GR_Ea05F	c 174	17.2	78.2	277	2	BB108898	BB108898
c 102	17.8	80.9	680	7	CO043288	UI-M-EHOP	CO043288	UI-M-EHOP	c 175	17.2	78.2	321	5	BX077413	BX077413
c 103	17.8	80.9	696	6	BY762798	BY762798	BY762798	BY762798	c 176	17.2	78.2	380	7	H57000	H57000
c 104	17.8	80.9	716	6	CB204291	AGENCOURT	CB204291	AGENCOURT	c 177	17.2	78.2	390	7	H56993	H56993
c 105	17.8	80.9	739	4	BG064637	H3025H05-	BG064637	H3025H05-	c 178	17.2	78.2	456	7	R54620	R54620
c 106	17.8	80.9	746	1	AA986905	ue16d12.x	AA986905	ue16d12.x	c 179	17.2	78.2	493	9	CE5656055	CE5656055
c 107	17.8	80.9	774	6	CB679920	OSUNEf04A	CB679920	OSUNEf04A	c 180	17.2	78.2	514	8	AQ149821	AQ149821
c 108	17.8	80.9	781	4	BM386120	UI-R-CNL-	BM386120	UI-R-CNL-	c 181	17.2	78.2	596	2	BE162184	BE162184
c 109	17.8	80.9	805	7	CK196621	FGAS100508	CK196621	FGAS100508	c 182	17.2	78.2	603	4	BJ195514	BJ195514
c 110	17.8	80.9	833	6	CB679919	OSUNEf04A	CB679919	OSUNEf04A	c 183	17.2	78.2	606	4	BJ604059	BJ604059
c 111	17.8	80.9	921	7	CF582656	AGENCOURT	CF582656	AGENCOURT	c 184	17.2	78.2	609	9	CE087212	CE087212
c 112	17.8	80.9	929	6	CB203977	AGENCOURT	CB203977	AGENCOURT	c 185	17.2	78.2	617	4	BM359006	BM359006
c 113	17.8	80.9	1090	9	CL089067	ISB1-12A2	CL089067	ISB1-12A2	c 186	17.2	78.2	627	4	BM3590667	BM3590667
c 114	17.4	79.1	253	1	AI470540	t190g11.x	AI470540	t190g11.x	c 187	17.2	78.2	629	2	BF270262	BF270262
c 115	17.4	79.1	275	5	BQ006680	UI-H-E11-	BQ006680	UI-H-E11-	c 188	17.2	78.2	630	4	BJ172664	BJ172664
c 116	17.4	79.1	280	6	CA412761	UI-H-E20-	CA412761	UI-H-E20-	c 189	17.2	78.2	637	4	BJ590311	BJ590311
c 117	17.4	79.1	283	5	BQ447114	UI-H-EU1-	BQ447114	UI-H-EU1-	c 190	17.2	78.2	638	4	BJ589567	BJ589567
c 118	17.4	79.1	289	6	CA419191	UI-H-E21-	CA419191	UI-H-E21-	c 191	17.2	78.2	663	5	BQ410318	BQ410318
c 119	17.4	79.1	294	5	BQ181527	UI-H-EU0-	BQ181527	UI-H-EU0-	c 192	17.2	78.2	673	4	BJ580428	BJ580428
c 120	17.4	79.1	344	7	CN480369	UI-H-EU0-	CN480369	UI-H-EU0-	c 193	17.2	78.2	677	6	CD375439	CD375439
c 121	17.4	79.1	345	5	BM988025	UI-H-DFO-	BM988025	UI-H-DFO-	c 194	17.2	78.2	685	4	BJ581569	BJ581569
c 122	17.4	79.1	366	7	CV372751	PM3-ET026	CV372751	PM3-ET026	c 195	17.2	78.2	701	4	BJ579010	BJ579010
c 123	17.4	79.1	381	7	CO093571	GR_Ea15D	CO093571	GR_Ea15D	c 196	17.2	78.2	702	4	BJ593096	BJ593096
c 124	17.4	79.1	384	7	CO599170	DG8-177p1	CO599170	DG8-177p1	c 197	17.2	78.2	711	4	BJ581862	BJ581862
c 125	17.4	79.1	386	5	BQ447790	UI-H-EU1-	BQ447790	UI-H-EU1-	c 198	17.2	78.2	717	5	BX625526	BX625526
c 126	17.4	79.1	390	2	AW207201	UI-H-B11-	AW207201	UI-H-B11-	c 199	17.2	78.2	720	9	CW533914	CW533914
c 127	17.4	79.1	398	2	BE218409	hv38c07.x	BE218409	hv38c07.x	c 200	17.2	78.2	725	4	BJ579893	BJ579893
c 128	17.4	79.1	402	2	BF196113	7n46d01.x	BF196113	7n46d01.x	c 201	17.2	78.2	727	4	BJ587322	BJ587322
c 129	17.4	79.1	411	1	AI684377	tc96f11.x	AI684377	tc96f11.x	c 202	17.2	78.2	731	4	BJ581518	BJ581518
c 130	17.4	79.1	413	5	BQ029011	UI-H-DFO-	BQ029011	UI-H-DFO-	c 203	17.2	78.2	738	4	BJ590545	BJ590545
c 131	17.4	79.1	416	5	BQ574629	UI-H-E21-	BQ574629	UI-H-E21-	c 204	17.2	78.2	739	4	BJ586364	BJ586364
c 132	17.4	79.1	467	1	AA937215	OK13f05..8	AA937215	OK13f05..8	c 205	17.2	78.2	743	7	CN651515	CN651515
c 133	17.4	79.1	470	5	BQ447743	UI-H-EU1-	BQ447743	UI-H-EU1-	c 206	17.2	78.2	754	7	CF517628	CF517628
c 134	17.4	79.1	471	5	BQ447819	UI-H-EU1-	BQ447819	UI-H-EU1-	c 207	17.2	78.2	757	7	CK773708	CK773708
c 135	17.4	79.1	471	5	BUE17686	UI-H-DFO-	BUE17686	UI-H-DFO-	c 208	17.2	78.2	770	4	BJ579007	BJ579007
c 136	17.4	79.1	472	5	BUE17986	UI-H-DFO-	BUE17986	UI-H-DFO-	c 209	17.2	78.2	775	4	BJ583282	BJ583282
c 137	17.4	79.1	473	5	BY457051	BY457051	BY457051	BY457051	c 210	17.2	78.2	775	4	BJ605785	BJ605785
c 138	17.4	79.1	477	5	BQ182663	UI-H-EU0-	BQ182663	UI-H-EU0-	c 211	17.2	78.2	780	4	BJ594593	BJ594593
c 139	17.4	79.1	490	6	BQ13462	UI-H-E20-	CA13462	UI-H-E20-	c 212	17.2	78.2	784	4	BJ590265	BJ590265
c 140	17.4	79.1	491	6	CA428195	UI-H-DFO-	CA428195	UI-H-DFO-	c 213	17.2	78.2	788	7	CF513224	CF513224
c 141	17.4	79.1	493	6	CA427568	UI-H-DFO-	CA427568	UI-H-DFO-	c 214	17.2	78.2	798	4	BJ583368	BJ583368
c 142	17.4	79.1	494	2	BE496100	WHE1261.D	BE496100	WHE1261.D	c 215	17.2	78.2	801	5	BP507093	BP507093
c 143	17.4	79.1	494	5	EX095415	EX095415	EX095415	EX095415	c 216	17.2	78.2	804	9	AG532109	AG532109
c 144	17.4	79.1	507	5	BUE17692	UI-H-DFO-	BUE17692	UI-H-DFO-	c 217	17.2	78.2	809	8	BZ284754	BZ284754
c 145	17.4	79.1	522	6	CA427653	UI-H-DFO-	CA427653	UI-H-DFO-	c 218	17.2	78.2	888	5	BU509171	BU509171
c 146	17.4	79.1	535	6	BQ447070	UI-H-EU1-	BQ447070	UI-H-EU1-	c 219	17.2	78.2	889	9	CG436661	CG436661
c 147	17.4	79.1	535	6	CA418441	UI-H-E21-	CA418441	UI-H-E21-	c 220	17.2	78.2	899	9	CW518344	CW518344
c 148	17.4	79.1	536	5	BQ46891	UI-H-EU1-	BQ46891	UI-H-EU1-	c 221	17.2	78.2	902	9	CL037878	CL037878
c 149	17.4	79.1	612	5	BQ573868	UI-H-E20-	BQ573868	UI-H-E20-	c 222	17.2	78.2	907	7	CF514216	CF514216
c 150	17.4	79.1	615	5	BQ447130	UI-H-EU1-	BQ447130	UI-H-EU1-	c 223	17.2	78.2	1010	9	CL099324	CL099324
c 151	17.4	79.1	624	4	B1236493	RE32552..5	B1236493	RE32552..5	c 224	17.2	78.2	1012	9	CL991530	CL991530
c 152	17.4	79.1	640	4	BJ209606	BJ209606	BJ209606	BJ209606	c 225	17.2	78.2	1109	6	CB182381	CB182381
c 153	17.4	79.1	653	5	BQ006783	UI-H-E11-	BQ006783	UI-H-E11-	c 226	17.2	78.2	1269	4	BM474622	BM474622
c 154	17.4	79.1	653	5	BQ445900	UI-H-EU1-	BQ445900	UI-H-EU1-	c 227	17.2	78.2	1269	4	BM474622	BM474622
c 155	17.4	79.1	661	5	BQ182753	UI-H-EU0-	BQ182753	UI-H-EU0-	c 228	17.2	78.2	1269	4	BM474622	BM474622
c 156	17.4	79.1	665	5	BUE16718	UI-H-DFO-	BUE16718	UI-H-DFO-	c 229	17.2	78.2	1269	4	BM474622	BM474622
c 157	17.4	79.1	673	4	BG926371	HNC55-1-H	BG926371	HNC55-1-H	c 230	17.2	78.2	1269	4	BM474622	BM474622
c 158	17.4	79.1	682	4	BJ217046	BJ217046	BJ217046	BJ217046	c 231	17.2	78.2	1269	4	BM474622	BM474622
c 159	17.4	79.1	685	5	BQ044947	UI-H-EU0-	BQ044947	UI-H-EU0-	c 232	17.2	78.2	1269	4	BM474622	BM474622
c 160	17.4	79.1	695	5	BUE17709	UI-H-DFO-	BUE17709	UI-H-DFO-	c 233	17.2	78.2	1269	4	BM474622	BM474622
c 161	17.4	79.1	703	5	BQ183342	UI-H-EU0-	BQ183342	UI-H-EU0-	c 234	17.2	78.2	1269	4	BM474622	BM474622
c 162	17.4	79.1	703	5	BQ717633	UI-H-E21-	BQ717633	UI-H-E21-	c 235	17.2	78.2	1269	4	BM474622	BM474622
c 163	17.4	79.1	710	5	BQ183771	UI-H-EU0-	BQ183771	UI-H-EU0-	c 236	17.2	78.2	1269	4	BM474622	BM474622
c 164	17.4	79.1	736	6	CD896159	G174.101P	CD896159	G174.101P	c 237	17.2	78.2	1269	4	BM474622	BM474622
c 165	17.4	79.1	772	5	BQ183620	UI-H-EU0-	BQ183620	UI-H-EU0-	c 238	17.2	78.2	1269	4	BM474622	BM474622
c 166	17.4	79.1	816	2	BF676759	602086413	BF676759	602086413	c 239	17.2	78.2	1269	4	BM474622	BM474622
c 167	17.4	79.1	874	5	BUI50273	AGENCOURT	BUI50273	AGENCOURT	c 240	17.2	78.2	1269	4	BM474622	BM474622
c 168	17.4	79.1	987	4	B1219997	602934668	B1219997	602934668	c 241	17.2	78.2	1269	4	BM474622	BM474622
c 169	17.4	79.1	1028	9	CNS02DMZ	Tetraodon	AL192644	Tetraodon	c 242	17.2	78.2	1269	4	BM474622	BM474622
c 170	17.4	79.1	1044	6	CB9997060	AGENCOURT	CB9997060	AGENCOURT	c 243	17.2	78.2	1269	4	BM474622	BM474622

390	16.4	74.5	441	5	BU694252	LL2GRID-3	463	16.4	74.5	872	9	AG526649	Mus muscu
391	16.4	74.5	446	1	AI296606	LP11024.5	C 464	16.4	74.5	880	9	CG267973	OG3CR05TV
392	16.4	74.5	448	8	B2324016	ic04b03.g	465	16.4	74.5	914	9	CG267960	OG3CR05TV
C 393	16.4	74.5	452	8	B58776	CIT-HSP-201	466	16.4	74.5	919	7	CK165414	FGAS04936
C 394	16.4	74.5	453	7	CO328673	EX2911139.	467	16.4	74.5	926	9	CG317707	OG0CM89TV
C 395	16.4	74.5	457	2	B8859849	BB859849	C 468	16.4	74.5	951	5	BUS26596	BUS26596
C 396	16.4	74.5	458	4	BG305457	F199111.x	C 469	16.4	74.5	979	7	CO648916	AGENCOURT
C 397	16.4	74.5	459	8	AQ228978	HS 2013.B	C 470	16.4	74.5	981	9	CNS002SD	AL098071
C 398	16.4	74.5	463	6	BY595474	BY595474	471	16.4	74.5	986	6	BY704853	Drosophill
C 399	16.4	74.5	493	2	BF165660	601777968	472	16.4	74.5	1001	5	BQ061118	AGENCOURT
400	16.4	74.5	494	4	AI609455	RH14240.5	C 473	16.4	74.5	1026	4	BM465164	AGENCOURT
C 401	16.4	74.5	502	2	AW344232	F176B06.x	474	16.4	74.5	1090	7	CF580404	AGENCOURT
C 402	16.4	74.5	513	1	AL917959	AL917959	475	16.4	74.5	1121	8	CC204350	CH261-95A
C 403	16.4	74.5	518	5	BX858805	BX858805	476	16.4	74.5	1152	4	BM454140	AGENCOURT
C 404	16.4	74.5	519	1	AI260821	LP04808.3	477	16.4	74.5	1168	4	BI549576	603192282
C 405	16.4	74.5	519	8	AZ234572	RPCI-23.5	478	16.4	74.5	1193	3	AY105439	Zea mays
406	16.4	74.5	527	2	BE225004	945042B02	C 479	16.4	74.5	1255	9	CU649095	CH213-214
407	16.4	74.5	539	1	AI259788	LP03280.5	480	16.4	74.5	1361	5	BUS38541	HUS38541
408	16.4	74.5	541	2	BB631820	BB631820	C 481	16.4	74.5	1462	3	BE962577	601655913
409	16.4	74.5	542	7	CO126126	GR-Eb09P	C 482	16.4	74.5	1465	3	AK039323	Mus muscu
410	16.4	74.5	555	4	BI294986	UI-E-DKO-	483	16.4	74.5	1749	3	AK011492	Mus muscu
411	16.4	74.5	558	7	CK658009	LP14485.5	484	16.4	74.5	2810	3	AK089943	Mus muscu
412	16.4	74.5	564	1	AI258850	LP02146.5	C 485	16.4	74.5	2823	3	AK019856	Mus muscu
413	16.4	74.5	566	1	AI259638	LP03104.5	486	16.4	74.5	3137	3	BC028885	Mus muscu
414	16.4	74.5	566	1	BX304328	BX304328	487	16.4	74.5	3345	3	AK004646	Mus muscu
415	16.4	74.5	568	7	CK662329	LP22279.5	488	16.4	74.5	4211	3	AK038344	Mus muscu
416	16.4	74.5	574	1	AI881514	606070803	C 489	16.2	73.6	146	8	AZ792657	2M0045H15
417	16.4	74.5	602	5	BQ256838	NISC-ko07	C 490	16.2	73.6	194	9	CG953577	MBEEN557R
C 418	16.4	74.5	603	1	AI293360	LP04902.3	C 491	16.2	73.6	201	8	AZ773327	IM0584J14
C 419	16.4	74.5	615	4	BG415717	HVSMK000	C 492	16.2	73.6	205	2	AW840670	RC4-CN000
C 420	16.4	74.5	616	8	AQ201028	RPCI11-46	493	16.2	73.6	244	2	AW673986	ba59d06.x
421	16.4	74.5	626	6	CB380204	iq72c12.y	494	16.2	73.6	278	2	BF355875	CM1-HT087
422	16.4	74.5	627	7	CO358925	DR-ATE-OR	C 495	16.2	73.6	282	2	BS566298	BB566298
C 423	16.4	74.5	634	4	BM290649	EST577183	C 496	16.2	73.6	283	8	AQ428773	CIT81-EI-
424	16.4	74.5	641	6	BY740372	BY740372	C 497	16.2	73.6	292	7	H35367	H35367
C 425	16.4	74.5	649	8	AZ114686	RPCI-23-4	498	16.2	73.6	306	4	BG313910	WHE2066.F
426	16.4	74.5	649	8	BZ020499	oeh06e01.	499	16.2	73.6	313	2	BB203639	BB203639
C 427	16.4	74.5	655	1	AZ992208	2M0276G16	500	16.2	73.6	321	2	BB219944	BB219944
428	16.4	74.5	662	1	AI293985	LP07261.5							
429	16.4	74.5	686	1	AI259294	LP02675.5							
C 430	16.4	74.5	687	5	BQ262308	f282f06.x							
C 431	16.4	74.5	692	9	CL912707	OA_ABA001							
432	16.4	74.5	697	6	BY710756	BY710756							
433	16.4	74.5	705	6	CB524491	UI-M-FY0-							
434	16.4	74.5	706	3	AK076125	Mus muscu							
435	16.4	74.5	711	7	CK661646	LP21275.5							
C 436	16.4	74.5	716	7	CK132079	LP13323.5							
C 437	16.4	74.5	723	4	BI820053	603037238							
C 438	16.4	74.5	723	6	CA348275	679542 NC							
439	16.4	74.5	727	7	CN837635	AGENCOURT							
C 440	16.4	74.5	731	5	BU374833	603813962							
C 441	16.4	74.5	732	6	CB518772	UI-M-GH0-							
C 442	16.4	74.5	732	6	CV527320	CS_GIL_04							
443	16.4	74.5	747	9	AG413802	Mus muscu							
444	16.4	74.5	756	7	CN525849	UI-M-HN0-							
445	16.4	74.5	756	6	CA348296	679566 NC							
446	16.4	74.5	758	6	CA348296	679566 NC							
447	16.4	74.5	767	9	AG495518	Mus muscu							
448	16.4	74.5	776	6	CB519795	UI-M-GH0-							
449	16.4	74.5	777	9	AG552143	Mus muscu							
C 450	16.4	74.5	797	6	CD782265	EST653626							
C 451	16.4	74.5	803	1	AA979940	MEST3-A4.							
C 452	16.4	74.5	805	7	CF243026	AGENCOURT							
C 453	16.4	74.5	809	7	CK634114	UI-M-HN0-							
C 454	16.4	74.5	815	8	CC336114	OGIBG33TV							
C 455	16.4	74.5	828	8	BZ529859	OGADG54TC							
C 456	16.4	74.5	843	8	AQ899297	HS 2013.B							
C 457	16.4	74.5	846	2	BF664502	602146260							
C 458	16.4	74.5	861	3	AK012826	Mus muscu							
459	16.4	74.5	866	5	BU959191	AGENCOURT							
C 460	16.4	74.5	866	9	CG317696	OG0CM89TV							
C 461	16.4	74.5	871	6	CD791572	EST622933							
462	16.4	74.5	871	9	AG458540	Mus muscu							

ALIGNMENTS

RESULT 1	CC249763	1240 bp	DNA	linear	GSS 13-MAY-2003
LOCUS	CH261-16B9_Sp6.1	CH261	Gallus gallus	genomic clone	CH261-16B9,
DEFINITION	genomic survey sequence.				
ACCESSION	CC249763				
VERSION	CC249763.1 GI:30586513				
KEYWORDS	GSS.				
SOURCE	Gallus gallus (chicken)				
ORGANISM	Gallus gallus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
AUTHORS	1. (bases 1 to 1240)				
TITLE	Kremizki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.				
JOURNAL	Gallus gallus BAC End Reads				
COMMENT	Unpublished (2003)				
CONTACT	Contact: Richard K. Wilson				
WASHTON	Genome Sequencing Center				
UNIVERSITY	Washington University School of Medicine				
EMAIL	Email: submissions@wustl.edu				
INERT	Inert Length: 182000 Std Error: 0.00				
SEQ	Seq primer: Sp6 ATTAGTGACACTATAG				
CLASS	Class: BAC ends				
HIGH	High quality sequence start: 20				
QUALITY	High quality sequence stop: 627.				
LOCATION	Location/Qualifiers				
FEATURES	1. .1240				
SOURCE	/organism="Gallus gallus"				

```

/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/cloned="CH261-16E9"
/sex="female"
/cell_line="UCD001, inbred 256"
/cloned_lib="CH261"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match      86.4%; Score 19; DB 8; Length 1240;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCTTG 19
|||||
Db 873 ATCCTTGCCCGCCCGCTTG 891
|||||

RESULT 2
CA366933      684 bp mRNA linear EST 06-NOV-2002
LOCUS         642829 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT37N02_D_G01 5',
DEFINITION    mRNA sequence.
ACCESSION     CA366933
VERSION       CA366933.1 GI:24677723
KEYWORDS      EST.
SOURCE        Oncorhynchus mykiss (rainbow trout)
ORGANISM      Oncorhynchus mykiss

REFERENCE
AUTHORS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 684)
Rexroad,C.B. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
TITLE         Sequence analysis of a rainbow trout cDNA library and creation of a
JOURNAL       gene index
COMMENT       Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@cccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.
Seq primer: AGCGGATACAAATTTACACAGGA.
FEATURES
source
1..684
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/cloned="1RT37N02_D_G01"
/tissue_type="pooled"
/lab_host="DH10B"
/cloned_lib="NCCWA 1RT"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."

ORIGIN
Query Match      85.5%; Score 18.8; DB 6; Length 684;
Best Local Similarity 90.9%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCTTGAAAT 22
|||||
Db 266 ATCCTTGCCCGCCCGCTTGACT 287
|||||

/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/cloned="CH261-16E9"
/sex="female"
/cell_line="UCD001, inbred 256"
/cloned_lib="CH261"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match      86.4%; Score 19; DB 8; Length 1240;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCTTG 19
|||||
Db 873 ATCCTTGCCCGCCCGCTTG 891
|||||

RESULT 3
CD497293      1158 bp mRNA linear EST 12-JUN-2003
LOCUS         CDA27-A09.y1d-s SHGC-CDA Gasterosteus aculeatus cDNA clone
DEFINITION    CDA27-A09 3', mRNA sequence.
ACCESSION     CD497293
VERSION       CD497293.1 GI:31424324
KEYWORDS      EST.
SOURCE        Gasterosteus aculeatus (three spined stickleback)
ORGANISM      Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1158)
Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 27
High quality sequence start: 43
High quality sequence stop: 561.
FEATURES
source
1..1158
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/cloned="CDA27-A09"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stages="adult"
/cloned_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-CDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN
Query Match      85.5%; Score 18.8; DB 6; Length 1158;
Best Local Similarity 90.9%; Pred. No. 9.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCTTGAAAT 22
|||||
Db 1136 ATCCTTGCCCGCCCGCTTTAAT 1157
|||||

/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/cloned="CH261-16E9"
/sex="female"
/cell_line="UCD001, inbred 256"
/cloned_lib="CH261"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match      85.5%; Score 18.8; DB 6; Length 684;
Best Local Similarity 90.9%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCTTGAAAT 22
|||||
Db 266 ATCCTTGCCCGCCCGCTTGACT 287
|||||

RESULT 4
A1972057/c     398 bp mRNA linear EST 27-OCT-1999
LOCUS         wv31h04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:2531191 3',
DEFINITION    mRNA sequence.
ACCESSION     A1972057
VERSION       A1972057.1 GI:5768883
KEYWORDS      EST.

```

```

SOURCE      Homo sapiens (human)
ORGANISM

REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 398)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 394.
FEATURES     Location/Qualifiers
            source
            1..398
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2531191"
            /tissue_type="fibrothema"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NCI CGAP Cv18"
            /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
            modified polylinker; Site 1: Not I; Site 2: Eco RI; let
            strand cDNA was primed with a Not I - oligo(dT) primer [5',
            TCTTACCAATCGAGTGGAGCGCGCGCGACATTTTTTTTTTTTTTTT 3'];
            double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not
            I and Eco RI sites of the modified pT7T3 vector. Library
            went through one round of normalization, and was
            constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match      83.6%; Score 18.4; DB 1; Length 398;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCCCCCGCCCTTGAAT 22
    |||||||||||||||
DB 313 CCGTCCCGCCCGCCCTTGAAT 294

RESULT 5
BF511913/c
LOCUS      BF511913      408 bp      mRNA      linear      EST 06-DEC-2000
DEFINITION UI-H-B14-apt-b-07-0-UI.s1 NCI CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088453 3', mRNA sequence.
ACCESSION  BF511913
VERSION     BF511913.1 GI:11595211
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 408)
AUTHORS    Katakis, I., et al.
TITLE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NCI-CGAP clone distribution
            information can be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 408.
FEATURES     Location/Qualifiers
            source
            1..408
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3088453"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="NCI CGAP_Sub8"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP_Sub8
            is a subtracted library derived from NCI CGAP_Sub5. The
            NCI CGAP_Sub8 library had 2.5 million recombinants. A
            single-stranded DNA preparation of NCI CGAP_Sub5 was used
            as a tracer in a subtractive hybridization with a driver
            comprising: a pool of clones from NCI CGAP_Sub5 (IMAGE
            clone ids 2732833-2737415, 3068040-3069191; 25% of the
            driver population), a pool of clones from NCI CGAP_Sub4
            (IMAGE clone ids 2723592-2729326; 25% of the driver
            population), NCI CGAP_Sub6 (pool AIF-AJU, IMAGE ids
            2728969-2733190; 25% of the driver population), and
            NCI CGAP_Sub7 (IMAGE ids 3069192-3072238,
            3081864-3084550; 25% of the driver population).
            Subtraction was performed as previously described
            (Bonaldo, Lennon & Soares (1996): Normalization and
            Subtraction: Two Approaches to Facilitate Gene Discovery.
            Genome Research 6, 791-806.
            TAG_SEQ=None found"
ORIGIN
Query Match      83.6%; Score 18.4; DB 2; Length 408;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCCCCCGCCCTTGAAT 22
    |||||||||||||||
DB 354 CCGTCCCGCCCGCCCTTGAAT 335

RESULT 6
BF430395/c
LOCUS      BF430395      436 bp      mRNA      linear      EST 29-NOV-2000
DEFINITION 933 MARC BSM Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BF430395
VERSION     BF430395.1 GI:11442492
KEYWORDS   EST.
SOURCE     Bos taurus (cow)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1 (bases 1 to 436)
AUTHORS    Warren, W.C., Tao, N., Allison, T., Wagner, S., Mathialagan, N.,
            Kata, S., Johnson, J., Smith, T.P.L. and Womack, J.
TITLE      A survey of genes transcribed in bovine skeletal muscle
JOURNAL    Unpublished (2000)
COMMENT    Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smiththe@mail.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: GGAAACAGCTATGACCATG
            BACKWARD: GTTITCCAGTCACGAC
            Seq primer: AATTAACCCCTCACTAAAGGG.
            Location/Qualifiers
            source
            1..436
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3088453"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="NCI CGAP_Sub8"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP_Sub8
            is a subtracted library derived from NCI CGAP_Sub5. The
            NCI CGAP_Sub8 library had 2.5 million recombinants. A
            single-stranded DNA preparation of NCI CGAP_Sub5 was used
            as a tracer in a subtractive hybridization with a driver
            comprising: a pool of clones from NCI CGAP_Sub5 (IMAGE
            clone ids 2732833-2737415, 3068040-3069191; 25% of the
            driver population), a pool of clones from NCI CGAP_Sub4
            (IMAGE clone ids 2723592-2729326; 25% of the driver
            population), NCI CGAP_Sub6 (pool AIF-AJU, IMAGE ids
            2728969-2733190; 25% of the driver population), and
            NCI CGAP_Sub7 (IMAGE ids 3069192-3072238,
            3081864-3084550; 25% of the driver population).
            Subtraction was performed as previously described
            (Bonaldo, Lennon & Soares (1996): Normalization and
            Subtraction: Two Approaches to Facilitate Gene Discovery.
            Genome Research 6, 791-806.
            TAG_SEQ=None found"
ORIGIN
Query Match      83.6%; Score 18.4; DB 2; Length 408;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCCCCCGCCCTTGAAT 22
    |||||||||||||||
DB 354 CCGTCCCGCCCGCCCTTGAAT 335

RESULT 6
BF430395/c
LOCUS      BF430395      436 bp      mRNA      linear      EST 29-NOV-2000
DEFINITION 933 MARC BSM Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BF430395
VERSION     BF430395.1 GI:11442492
KEYWORDS   EST.
SOURCE     Bos taurus (cow)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1 (bases 1 to 436)
AUTHORS    Warren, W.C., Tao, N., Allison, T., Wagner, S., Mathialagan, N.,
            Kata, S., Johnson, J., Smith, T.P.L. and Womack, J.
TITLE      A survey of genes transcribed in bovine skeletal muscle
JOURNAL    Unpublished (2000)
COMMENT    Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smiththe@mail.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: GGAAACAGCTATGACCATG
            BACKWARD: GTTITCCAGTCACGAC
            Seq primer: AATTAACCCCTCACTAAAGGG.
            Location/Qualifiers
            source
            1..436
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3088453"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="NCI CGAP_Sub8"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP_Sub8
            is a subtracted library derived from NCI CGAP_Sub5. The
            NCI CGAP_Sub8 library had 2.5 million recombinants. A
            single-stranded DNA preparation of NCI CGAP_Sub5 was used
            as a tracer in a subtractive hybridization with a driver
            comprising: a pool of clones from NCI CGAP_Sub5 (IMAGE
            clone ids 2732833-2737415, 3068040-3069191; 25% of the
            driver population), a pool of clones from NCI CGAP_Sub4
            (IMAGE clone ids 2723592-2729326; 25% of the driver
            population), NCI CGAP_Sub6 (pool AIF-AJU, IMAGE ids
            2728969-2733190; 25% of the driver population), and
            NCI CGAP_Sub7 (IMAGE ids 3069192-3072238,
            3081864-3084550; 25% of the driver population).
            Subtraction was performed as previously described
            (Bonaldo, Lennon & Soares (1996): Normalization and
            Subtraction: Two Approaches to Facilitate Gene Discovery.
            Genome Research 6, 791-806.
            TAG_SEQ=None found"
ORIGIN
Query Match      83.6%; Score 18.4; DB 2; Length 408;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCCCCCGCCCTTGAAT 22
    |||||||||||||||
DB 354 CCGTCCCGCCCGCCCTTGAAT 335

```

```

/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Skeletal muscle"
/lab_host="XLOLR"
/clone_lib="MARC BSM"
/notes="Vector: Uni-ZAP XR, Site 1: EcoRI; Site 2: XhoI;
Library obtained from Strategene, catalog #937721. Library
made from skeletal muscle of a two year old Holstein cow."

ORIGIN
Query Match      83.6%; Score 18.4; DB 2; Length 436;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAAT 22
|||||
Db 75 CCTTGGCCCGCCCTTGAAT 56

RESULT 7
BH828773/c
LOCUS BH828773 440 bp DNA linear GSS 20-MAY-2002
DEFINITION BACP28-O13.z Pristionchus pacificus BAC ends Pristionchus
pacificus genomic, genomic survey sequence.
ACCESSION BH828773
VERSION BH828773.1 GI:21023647
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 440)
AUTHORS Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R.,
Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A.,
Jesse,T., Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.
TITLE A BAC-based genetic linkage map of the nematode Pristionchus
pacificus
JOURNAL Unpublished (2002)
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..440
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="var. California"
/db_xref="taxon:54126"
/clone_lib="Pristionchus pacificus BAC ends"

ORIGIN
Query Match      83.6%; Score 18.4; DB 8; Length 440;
Best Local Similarity 95.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 21
|||||
Db 32 TCCTTGGCCCGCCCTTGAAT 13

RESULT 8
BF507848/c
LOCUS BF507848 473 bp mRNA linear EST 06-DEC-2000
DEFINITION UI-H-B14-apv-h-01-0-UI-s1 NCI CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088753 3', mRNA sequence.
ACCESSION BF507848
VERSION BF507848.1 GI:11591146
KEYWORDS EST.

```

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
Location/Qualifiers
1..473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3088753"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Sub8"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub8
is a subtraced library derived from NCI CGAP Sub5. The
NCI CGAP Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub5 was used
as a tracer in a subtractive hybridization with a driver
comprising: a pool of clones from NCI CGAP Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population), a pool of clones from NCI CGAP Sub4
(IMAGE clone ids 2723592-2729326; 25% of the driver
population), NCI CGAP Sub6 (pool AIF-AJU, IMAGE ids
2728969-2733190; 25% of the driver population), and
NCI CGAP Sub7 (IMAGE ids 3069192-3072238,
3081864-3084550; 25% of the driver population).
Subtraction was performed as previously described
(Bonaldo, Lennon & Soares (1996): Normalization and
Subtraction: two Approaches to Facilitate Gene Discovery.
Genome Research 6, 791-806.
TAG_SEQ=None found"

ORIGIN
Query Match      83.6%; Score 18.4; DB 2; Length 473;
Best Local Similarity 95.0%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAAT 22
|||||
Db 354 CGTGGCCCGCCCTTGAAT 335

RESULT 9
A1380413/c
LOCUS A1380413 478 bp mRNA linear EST 30-MAR-1999
DEFINITION tf96c06.x1 NCI CGAP_CL11 Homo sapiens cDNA clone IMAGE:2107114 3',
mRNA sequence.
ACCESSION A1380413
VERSION A1380413.1 GI:4190266
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html
Insert length: 872 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 460.

FEATURES source
1. 478 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2107114"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/clone_lib="NCI_CGAP CLL1"
/note="Vector: pVT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGAGCGCGCATTTCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pVT3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match 83.6%; Score 18.4; DB 1; Length 478;
Best Local Similarity 95.0%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAAT 22
Db 337 CCGTGGCCCGCCCTTGAAT 318

RESULT 10
CA397297/c
LOCUS CA397297 638 bp mRNA linear EST 06-NOV-2002
DEFINITION cs99d06.y1 Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs99d06 5', mRNA sequence.

ACCESSION CA397297
VERSION CA397297.1 GI:24734475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 638)
Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: Over 6000 non-redundant transcripts, novel genes and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
22103460
22103460
MEDLINE
PUBMED
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 89 row: d column: 06
Seq primer: M13RP1 reverse primer (ABI).

FEATURES source
1. 638 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="cs99d06"
/tissue_type="RPE/choroid"
/dev stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs"
/note="Organ: Eye; Vector: pCMVSPORT6; Two different donor eyes (75-80 years old) yielded approximately 600 mg of dissected RPE/choroid tissue. This in turn yielded 340 ug of total RNA and 7 ug of mRNA. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies (Rockville, MD; now part of Invitrogen Corp), essentially following the protocols of the SuperScript Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/>. The library code designation was cs. For this library, cDNA inserts were cloned into the NotI/MluI sites of the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN
Query Match 83.6%; Score 18.4; DB 6; Length 638;
Best Local Similarity 95.0%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAAT 22
Db 353 CCGTGGCCCGCCCTTGAAT 334

RESULT 11
AQ939474/c
LOCUS AQ939474 722 bp DNA linear GSS 23-AUG-2000
DEFINITION NR1-180R Human NotI clones Homo sapiens genomic, genomic survey sequence.

ACCESSION AQ939474
VERSION AQ939474.1 GI:7215952
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 722)
Zabarovsky,E.R., Gizatullin,R., Podowski,R.M., Zabarovska,V.V., Xie,L., Muravenko,O.V., Kozyrev,S., Petrenko,L., Skobeleva,N., Li,J., Protopopov,A., Kashuba,V., Ernberg,I., Winberg,G. and Wahlestedt,C.
Noti clones in the analysis of the human genome
Nucleic Acids Res. 28 (7), 1635-1639 (2000)
20175728
10710430
PUBMED
COMMENT Contact: Podowski RM
Center for Genomics Research
Karolinska Institute
17177 Stockholm, Sweden
Tel: +46-8-728-6372
Fax: +46-8-337983
Email: Raf.Podowski@cgr.ki.se
Class: Noti site.

FEATURES source
1. 722 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Human Noti clones"

ORIGIN
Query Match 83.6%; Score 18.4; DB 8; Length 722;

Best Local Similarity 95.0%; Pred. No. 1.4e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAA 22
Db 354 CCGTGGCCCGCCCTTGAA 335

RESULT 12
BE957426/c
LOCUS
DEFINITION BE957426 918 bp mRNA linear EST 14-DEC-2000
601653569R2 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838553 3',
mRNA sequence.
ACCESSION BE957426 GI:11773878
VERSION BE957426
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 918)
COMMENT NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Oct 3, 2000 this sequence version replaced gi:10568235.
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LMG525 row: i column: 18
High quality sequence start: 34
High quality sequence stop: 79.

FEATURES

source
1..918
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3838553"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 55"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgcttcggcc); Site_2: SfiI
(ggcgcttcggcc); Double-stranded cDNA was prepared from
cell line RNA. 5' and 3' adaptors were used in cloning as
follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and
3' adaptor sequence:
5'-ATTCAGAGCCGCGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.65 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

ORIGIN

Query Match 83.6%; Score 18.4; DB 2; Length 918;
Best Local Similarity 95.0%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCTTGGCCCGCCCTTGAA 21
Db 446 TCCTTGGCCCGCCCTTGAA 427

RESULT 13
AL567601/c

LOCUS
DEFINITION AL567601 1139 bp mRNA linear EST 05-APR-2004
AL567601 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone

CSODF030YH23 3-PRIME, mRNA sequence.

AL567601
AL567601.3 GI:46234082
EST.

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1139)

Li.W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 16, 2001 this sequence version replaced gi:31290473.

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

This sequence belongs to sequence cluster 2863.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?c=CSODF030CD12NP1&c=2863.f.

Location/Qualifiers

1..1139

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODF030YH23"

/tissue_type="FETAL BRAIN"

/dev_stage="fetal"

/clone_lib="Homo sapiens FETAL BRAIN"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 81.8%; Score 18; DB 1; Length 1139;
Best Local Similarity 75.0%; Pred. No. 2.2e+03;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCCTTGGCCCGCCCTTGAA 21

Db 1042 TCCTTGGCCCGCCCTTGAA 1023

RESULT 14

BM208321/c

LOCUS 269 bp mRNA linear EST 08-JUN-2003

DEFINITION C0625F08-3 NTA Mouse Trophoblast Stem Cell cDNA Library (Long) Mus
musculus cDNA clone NIA:C0625F08 IMAGE:30022627 3', mRNA sequence.

ACCESSION BM208321

VERSION BM208321.2 GI:31532256

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 269)

Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.

Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method

Genome Res. 11 (9), 1553-1558 (2001)

JOURNAL

MEDLINE 21429098

PUBMED 11544199

COMMENT

On Dec 14, 2001 this sequence version replaced gi:17764465.

Other ESTs: C0625F08-5N

Contact: Dawood B. Dudekula


```

Db      13 TCCTTCCCGCCGCTTGCAT 33

RESULT 16
BM222541/c
LOCUS   BM222541.2 331 bp mRNA linear EST 07-JUN-2003
DEFINITION K0122G08-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-)
            CDNA Library (long) Mus musculus cDNA clone NIA:K0122G08
IMAGE:30041551 3', mRNA sequence.

ACCESSION BM222541.2 GI:31480822
VERSION   1
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE 1 (bases 1 to 331)
AUTHORS   Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE   21429098
PUBMED    11544199
COMMENT   On Dec 14, 2001 this sequence version replaced gi:17783184.
            Other ESTs: K0122G08-5N
            Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@igsun.grc.nia.nih.gov
            Plate: K0122 row: G column: 08
            Seq primer: -21M13 Forward
            High quality sequence stop: 331
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..331
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6NCr"
                     /db_xref="niaEST:K0122G08-3"
                     /db_xref="taxon:10090"
                     /clone="NIA:K0122G08 IMAGE:30041551"
                     /tissue_type="Hematopoietic Stem Cell"
                     (lin-/c-Kit+/Sca-1-)
                     /dev_stage="Age approx. 10 weeks old"
                     /lab_host="PH108"
                     /clone_lib="NIA Mouse Hematopoietic Stem Cell"
                     (lin-/c-Kit+/Sca-1-) cDNA Library (long)
                     /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     obtained from Drs. Dennis Taub, Dan Longo (National
                     Institute on Aging, USA), Jonathan Keller (National Cancer
                     Institute, USA). Double-stranded cDNAs were synthesized
                     with an Oligo(dT) primer [Invitrogen:
                     5'-pGACAGTGTCTAGTCGAGCGCGCCCTTTTCTTTT-3'] from
                     2.4 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Loner-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.2 kb. The library was constructed
                     by Yulan Piao (NIA)."

```

ORIGIN

```

Query Match      80.9%; Score 17.8; DB 4; Length 331;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTCCCGCCGCTTGCAT 22
    ||||| ||| ||||| |||||
Db 134 TCCTTCCCGCTTGCAT 114

```

RESULT 17

```

BY393360      337 bp mRNA linear EST 12-DEC-2002
BY393360 RIKEN full-length enriched, 14 days pregnant adult female
placenta Mus musculus cDNA clone I530022M10 3', mRNA sequence.

ACCESSION BY393360
VERSION   1
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE 1 (bases 1 to 337)
AUTHORS   Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
            Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
            Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
            Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
            Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
            Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
            Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
            Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
            Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
            Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
            Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
            Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
            Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
            Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
            Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
            Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
            Sadelain, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
            Sultan, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
            Vardaro, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
            Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
            Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
            Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
            Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
            Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
            Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
            Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
            Rogers, J., Birney, E. and Hayashizaki, Y.

```

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1..337
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I530022M10"
 /sex="female"
 /tissue_type="placenta"
 /dev_stage="14 days pregnant adult"
 /clone_lib="RIKEN full-length enriched, 14 days pregnant adult female placenta"

ORIGIN

Query Match 80.9%; Score 17.8; DB 5; Length 337;
 Best Local Similarity 90.5%; Pred. No. 2.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22
 ||||| ||| ||||| |||||
 Db 46 TCCTTCCCTGCGCCCTTGAAT 66

RESULT 18
 AA959728/c
 LOCUS
 DEFINITION
 IMAGE:1247628 3', mRNA sequence.
 ACCESSION
 AA959728
 VERSION
 AA959728.1 GI:3125628
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 338)
 REFERENCE
 AUTHORS
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Seftoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished (1996)
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:661316
 Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES

source

Location/Qualifiers

1..338
 /organism="Mus musculus"
 /mol_type="mRNA"

/db_xref="taxon:10090"
 /clone="IMAGE:1247628"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DHI0B"

/clone_lib="Soares mammary_gland_NMLMG"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 338;
 Best Local Similarity 90.5%; Pred. No. 2.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22
 ||||| ||| ||||| |||||
 Db 336 TCCTTCCCTGCGCCCTTGAAT 316

RESULT 19
 AA959831/c
 LOCUS
 DEFINITION
 IMAGE:1247676 3', mRNA sequence.
 ACCESSION
 AA959831
 VERSION
 AA959831.1 GI:3125731
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 339)
 REFERENCE
 AUTHORS
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Seftoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished (1996)
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:661364
 Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES
 source
 1..339
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:1247676"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DHI0B"

/clone_lib="Soares mammary_gland_NMLMG"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library

```

SOURCE ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Angen EST Program
TITLE
Angen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00002 row: a column: 2.
FEATURES
source
1..348
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrpc1-00002-a2"
/tissue_type="adrenal"
/clone_lib="nrpc1 (10212)"
/note="Vector: pX; Site 1: SalI; Site 2: NotI;
pneochromocytoma adrenal cells"
ORIGIN
Query Match 80.9%; Score 17.8; DB 6; Length 348;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 100 TCCTTCCCTGCCCTTGAAT 120
RESULT 22
AA032751
LOCUS
DEFINITION
mi35a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:465482 5', mRNA sequence.
ACCESSION
AA032751
VERSION
AA032751.1 GI:1504162
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 352)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279298
Seq primer: -28M13 rev2 from Amersham.
FEATURES
source
1..352
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465482"

```

```

SOURCE ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Angen EST Program
TITLE
Angen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00002 row: a column: 2.
FEATURES
source
1..348
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrpc1-00002-a2"
/tissue_type="adrenal"
/clone_lib="nrpc1 (10212)"
/note="Vector: pX; Site 1: SalI; Site 2: NotI;
pneochromocytoma adrenal cells"
ORIGIN
Query Match 80.9%; Score 17.8; DB 6; Length 348;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 100 TCCTTCCCTGCCCTTGAAT 120
RESULT 22
AA032751
LOCUS
DEFINITION
mi35a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:465482 5', mRNA sequence.
ACCESSION
AA032751
VERSION
AA032751.1 GI:1504162
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 352)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279298
Seq primer: -28M13 rev2 from Amersham.
FEATURES
source
1..352
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465482"

```

```

SOURCE ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Angen EST Program
TITLE
Angen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00002 row: a column: 2.
FEATURES
source
1..348
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrpc1-00002-a2"
/tissue_type="adrenal"
/clone_lib="nrpc1 (10212)"
/note="Vector: pX; Site 1: SalI; Site 2: NotI;
pneochromocytoma adrenal cells"
ORIGIN
Query Match 80.9%; Score 17.8; DB 6; Length 348;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 100 TCCTTCCCTGCCCTTGAAT 120
RESULT 22
AA032751
LOCUS
DEFINITION
mi35a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:465482 5', mRNA sequence.
ACCESSION
AA032751
VERSION
AA032751.1 GI:1504162
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 352)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279298
Seq primer: -28M13 rev2 from Amersham.
FEATURES
source
1..352
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465482"

```

```

SOURCE ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Angen EST Program
TITLE
Angen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00002 row: a column: 2.
FEATURES
source
1..348
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrpc1-00002-a2"
/tissue_type="adrenal"
/clone_lib="nrpc1 (10212)"
/note="Vector: pX; Site 1: SalI; Site 2: NotI;
pneochromocytoma adrenal cells"
ORIGIN
Query Match 80.9%; Score 17.8; DB 6; Length 348;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 100 TCCTTCCCTGCCCTTGAAT 120
RESULT 22
AA032751
LOCUS
DEFINITION
mi35a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:465482 5', mRNA sequence.
ACCESSION
AA032751
VERSION
AA032751.1 GI:1504162
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 352)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279298
Seq primer: -28M13 rev2 from Amersham.
FEATURES
source
1..352
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465482"

```

/sex="unknown"
 /tissue type="embryo"
 /dev stage="13.5-14.5dpc total fetus"
 /lab host="DH10B"
 /clone_lib="Soares mouse embryo NBM13.5 14.5"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTCGAGCGCGCGGAATTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 352;
 Best Local Similarity 90.5%; Pred. No. 2.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22

Db 29 TCCTTCCCGCCCTTGAAT 49

RESULT 23

BY400112 357 bp mRNA linear EST 12-DEC-2002
 LOCUS BY400112 RIKEN full-length enriched, TIB-55 BB88 Mus musculus cDNA
 clone 1730028M19 3', mRNA sequence.

ACCESSION BY400112.1 GI:26229680

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 357)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Dalla, E., Dragani, T.A.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, D., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Harai, A., Hashizume, M., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 12466851

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

1..357
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /clone="I730028M19"
 /cell_line="TIB-55 BB88"
 /clone_lib="RIKEN full-length enriched, TIB-55 BB88"

ORIGIN

Query Match 80.9%; Score 17.8; DB 5; Length 357;
 Best Local Similarity 90.5%; Pred. No. 2.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22

Db 63 TCCTTCCCGCCCTTGAAT 83

RESULT 24

BY467320 361 bp mRNA linear EST 31-DEC-2002
 LOCUS BY467320 RIKEN full-length enriched, melanocyte Mus musculus cDNA
 clone G270027101 3', mRNA sequence.

ACCESSION BY467320

VERSION BY467320.1 GI:26801699

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 361)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

```

1..366
/organism="Mus musculus"
/mol_type="mRNA"
/strain="DBA/2"
/db_xref="taxon:10090"
/clone="I730087D07"
/cell_line="CRL-1722 L5178Y-R"
/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

```

ORIGIN

```

Query Match      80.9%; Score 17.8; DB 5; Length 366;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 2 TCCTTGCCCGCCCTTGAAT 22

Db 77 TCCTTCCCGTGCCTTGAAT 97

RESULT 26

BY405025

LOCUS

```

BY405025 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus
cDNA clone I730052E05 3', mRNA sequence.

```

BY405025

EST.

GI:26634593

SOURCE

Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 370)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaio, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hune, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Mateuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bratt, D., Brusic, V., Chochoa, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konegava, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sultana, R., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354583

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

```

1..370
/organism="Mus musculus"
/mol_type="mRNA"
/strain="DBA/2"
/db_xref="taxon:10090"
/clone="I730052E05"
/cell_line="CRL-1722 L5178Y-R"
/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

```

ORIGIN

```

Query Match      80.9%; Score 17.8; DB 5; Length 370;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 2 TCCTTGCCCGCCCTTGAAT 22

Db 79 TCCTTCCCGTGCCTTGAAT 99

RESULT 27

BY681852

LOCUS

```

BY681852 RIKEN full-length enriched, 12.5 days embryo Rathke's
pouches Mus musculus cDNA clone K920039N07 3', mRNA sequence.

```

BY681852

EST.

VERSION

SOURCE

ORGANISM

REFERENCE

AUTHORS

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 376)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaio, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hune, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Mateuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bratt, D., Brusic, V., Chochoa, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konegava, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sultana, R., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, S., Gissi, C., Godzik, J., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perteau, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
22354683
PUBMED
12466851
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

Location/Qualifiers
1. .376
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="K920039N07"
/tissue_type="Rathke's pouches"

FEATURES
source

ORIGIN
Query Match 80.9%; Score 17.8; DB 1; Length 379;

/dev stage="12.5 days embryo"
/clone_lib="RIKEN full-length enriched, 12.5 days embryo Rathke's pouches"

ORIGIN

Query Match 80.9%; Score 17.8; DB 6; Length 376;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22

Db 73 TCCTTGGCCCGCCCTTGAAT 93

RESULT 28
AA260010

LOCUS
AA260010 379 bp mRNA linear EST 18-MAR-1997
IMAGE:746458 5', mRNA sequence.

DEFINITION
AA260010

ACCESSION
AA260010

VERSION
AA260010.1 GI:1896496

KEYWORDS
EST.

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 379)

AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
The WashU-HMI Mouse EST Project

JOURNAL
Unpublished (1996)

COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:455442

Putative full length read
vector to vector length is 542
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 338.

Location/Qualifiers
1. .379
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:746458"
/sex="unknown"
/tissue_type="fetus"
/dev stage="12.5dpc total fetus"
/lab_host="DH10B"

clone_lib="Soares mouse 3NWE12 5"
/notes="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCTATTTTTTTTTTTTTTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 80.9%; Score 17.8; DB 1; Length 379;

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES

Location/Qualifiers
1..383
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G93004H13"
/tissue_type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
/clone_lib="RCB-0527 Jyg-MC(B) cDNA"
RCB-0527 Jyg-MC(B) cDNA

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 383;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22
||||| ||||| ||||| ||||| |||||
Db 96 TCCTTGCCCGCCCTTGAAT 116

RESULT 31
BM208863/c
LOCUS
DEFINITION
BM208863
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
21429098
11544199
On Dec 14, 2001 this sequence version replaced gi:17765170.
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
Plate: C0636 row: B column: 05
Seq primer: -21M13 Forward
High quality sequence stop: 383
POLYA=Yes.

FEATURES

Location/Qualifiers
1..383
/organism="Mus musculus"
/mol_type="mRNA"
/strains="B5/EGFP transgenic ICR mice"
/db_xref="taxon:10090"

/clone="NTA:C0636B05 IMAGE:30023632"
/tissue_type="Trophoblast stem cell"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_lib="NTA Mouse Trophoblast Stem Cell cDNA Library
(Long)"
/note="vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NTA), Intramural Research
Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Janet Rossant and Pilo Kunath (Samuel
Lunenfeld Research Institute, Canada). Double-stranded
cDNAs were synthesized with an Oligo(dT) primer
[Invitrogen]:
5'-TGACTAGTCTAGATCGGCGCGCCCTTTT-3' from
4 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer SalI-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.6 kb. The library was constructed
by Yulan Piao (NTA)."

ORIGIN

Query Match 80.9%; Score 17.8; DB 4; Length 383;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAAT 22
||||| ||||| ||||| ||||| |||||
Db 293 TCCTTGCCCGCCCTTGAAT 273

RESULT 32
AA763480
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA763480 384 bp mRNA linear EST 27-JAN-1998
vw54907.t1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1247676 5', mRNA sequence.

AA763480
AA763480.1 GI:2813227
EST.

Mus musculus (house mouse)
Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 384)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LInL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:661364

Putative full length read
vector to vector length is 414

```

Seq primer: -28ml3 rev2 ET from Amerham.
FEATURES
    source
        1..384
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:1247676"
            /sex="female (lactating)"
            /tissue_type="mammary gland"
            /lab_host="DH10B"
            /clone_lib="Soares mammary gland NMLMG"
            /note="Vector: pRT3B-Pac (Pharmacia) with a modified
            polylinker; 1st strand cDNA was prepared from mammary
            gland tissue from a lactating female, and was then primed
            with a Not I - oligo(dT) primer. Double-stranded cDNA was
            ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the
            modified pRT3 vector. Library is normalized. Library
            was constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
    Query Match      80.9%; Score 17.8; DB 1; Length 384;
    Best Local Similarity 90.5%; Pred. No. 2.4e+03;
    Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  2 TCCTTGCCCGCCGCTTGAAT 22
    ||||| ||||| ||||| |||||
Db   36 TCCTTCCCTGCCCTTGAAT 56

RESULT 33
LOCUS BY677468 385 bp mRNA linear EST 16-DEC-2002
DEFINITION BY677468 RIKEN full-length enriched, 12.5 days embryo Rathke's
ACCESSION pouches Mus musculus cDNA clone K920014D19 3', mRNA sequence.
VERSION BY677468
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baidarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bratt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gassterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustinich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Nunata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Vardaro, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Satoh, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
ANALYSIS Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
NATURE 420, 563-573 (2002)

```

```

MEDLINE
PUBMED
COMMENT
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequences analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michelle Brinkmeier and Sally Camper (
Dept. Human Genetics University of Michigan Medical School 4301
MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA )
whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
    source
        1..385
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="CD-1"
            /db_xref="taxon:10090"
            /clone="K920014D19"
            /tissue_type="Rathke's pouches"
            /dev_stages="12.5 days embryo"
            /clone_lib="RIKEN full-length enriched, 12.5 days embryo
            Rathke's pouches"
ORIGIN
    Query Match      80.9%; Score 17.8; DB 6; Length 385;
    Best Local Similarity 90.5%; Pred. No. 2.4e+03;
    Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  2 TCCTTGCCCGCCGCTTGAAT 22
    ||||| ||||| ||||| |||||
Db   83 TCCTTCCCTGCCCTTGAAT 103

RESULT 34
LOCUS AI036434 387 bp mRNA linear EST 26-JUN-1998
DEFINITION ub64b04.r1 Soares mammary_gland_NMLMG Mus musculus cDNA clone
ACCESSION IMAGE:1382479 5', mRNA sequence.
VERSION AI036434
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 387)

```

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, K., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:904947
Seq primer: -28ml3 rev2 ET from Amerham.

FEATURES

source	1..387
Location/Qualifiers	
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
	/clone="IMAGE:1382479"
	/sex="female (lactating)"
	/tissue_type="mammary gland"
	/lab_host="DH10B"
	/clone_lib="Soares mammary gland NMLMG"
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match	80.9%;	Score 17.8;	DB 1;	Length 387;
Best Local Similarity	90.5%;	Pred. No. 2.4e+03;		
Matches	19;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;

Oy 2 TCCTTGCCCGCCCTTGAAT 22
|||||
Db 37 TCCTTCCCGTGCCTTGAAT 57
|||||

RESULT 35

BY409902

LOCUS BY409902 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus cDNA clone I730078A07 3', mRNA sequence.

DEFINITION BY409902

ACCESSION BY409902.1 GI:26639488

VERSION BY409902.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 387)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikiado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konggaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Nunata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M., Sadelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-5222
Fax: 81-45-503-5216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source	Location/Qualifiers
	1..387
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="DBA/2"
	/db_xref="taxon:10090"
	/clone="I730078A07"
	/cell_line="CRL-1722 L5178Y-R"
	/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

ORIGIN

Query Match	80.9%;	Score 17.8;	DB 5;	Length 387;
Best Local Similarity	90.5%;	Pred. No. 2.4e+03;		
Matches	19;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;

Oy 2 TCCTTGCCCGCCCTTGAAT 22
|||||
Db 101 TCCTTCCCGTGCCTTGAAT 121
|||||

RESULT 36

BY701593

BY701593 387 bp mRNA linear EST 16-DEC-2002
 BY701593 RIKEN full-length enriched, synovial fibroblasts Mus
 musculus cDNA clone L030003C11 3', mRNA sequence.
 BY701593
 BY701593.1 GI:27112668
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 387)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikiado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuoka, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konegaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Tissues were provided by Vassilis Aidinis (Biomedical Sciences
 Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
 Fleming street 16672 Vari, Greece) whose assistance we gratefully
 acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 Location/Qualifiers
 1..387
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="L030003C11"
 /cell_type="synovial fibroblasts"
 /clone_lib="RIKEN full-length enriched, synovial
 fibroblasts"
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 6; Length 387;
 Best Local Similarity 90.5%; Pred. NO. 2.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TCCTTGCCCGCCCGCTTGAAT 22
 ||||| ||| ||||| |||||
 Db 95 TCCTTCCCGCTTGAAT 115
 RESULT 37
 BY440840
 LOCUS
 DEFINITION
 BY440840 RIKEN full-length enriched, 16 days embryo kidney Mus
 musculus cDNA clone I920172K22 3', mRNA sequence.
 BY440840
 BY440840.1 GI:26725832
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 392)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikiado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuoka, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konegaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES
source
1. .392
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I920172K22"
/tissue_type="kidney"
/dev_stage="16 days embryo"
/clone_lib="RIKEN full-length enriched, 16 days embryo kidney"
ORIGIN
Query Match 80.9%; Score 17.8; DB 5; Length 392;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCCGCCCTTGAAT 22
|||||
DB 99 TCCTTCCCGTCCCGCTTGAAT 119
RESULT 38
BB794167
LOCUS
BB794167 RIKEN full-length enriched, brain CRL-1443 BC3H1 cDNA Mus
DEFINITION
musculus cDNA clone G430149C16 3', mRNA sequence.
ACCESSION
BB794167.1 GI:16963786
VERSION
BB794167.1
KEYWORDS
EST.
MUS musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 396)
REFERENCE
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hironoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yaeunishi, A.,

Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Kira, A., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. .396
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"
/clone="G430149C16"
/tissue_type="brain"
/cell_line="CRL-1443 BC3H1"
/clone_lib="RIKEN full-length enriched, brain CRL-1443 BC3H1 cDNA"
ORIGIN
Query Match 80.9%; Score 17.8; DB 2; Length 396;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCCGCCCTTGAAT 22
|||||
DB 110 TCCTTCCCGTCCCGCTTGAAT 130
RESULT 39
W88143
LOCUS
W88143.1
DEFINITION
mf69c02.r1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA
clone IMAGE:419522 5' similar to TR:G1203895 G1203895 ISOPEPTIDASE T. [1] ;, mRNA sequence.
ACCESSION
W88143
VERSION
W88143.1 GI:1542518
KEYWORDS
EST.
MUS musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 401)
REFERENCE
Matra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
JOURNAL
COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1402230.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:254074
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 359.

FEATURES
source

1..401
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:419522"
/sex="unknown"
/tissue_type="embryo"
/dev_stages="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NBM13.5 14.5"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pRT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 7; Length 401;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGGCCCGCCCTTGAAT 22
||||| ||| ||||| |||||
Db 337 TCCTTCCCTCGCCCTTGAAT 357

RESULT 40
BM119162/c
LOCUS
DEFINITION
BM119162 412 bp mRNA linear EST 01-FEB-2002
L0921F05-3 NTA Mouse Newborn Kidney cDNA Library (Long) Mus
musculus cDNA clone L0921F05 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 412)
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
and Ko, M.S.H.
Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library
Unpublished (2001)
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0921 row: F column: 05

TITLE
JOURNAL
COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1402230.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:254074
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 359.

FEATURES
source

1..401
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:419522"
/sex="unknown"
/tissue_type="embryo"
/dev_stages="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NBM13.5 14.5"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pRT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 7; Length 401;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGGCCCGCCCTTGAAT 22
||||| ||| ||||| |||||
Db 337 TCCTTCCCTCGCCCTTGAAT 357

RESULT 40
BM119162/c
LOCUS
DEFINITION
BM119162 412 bp mRNA linear EST 01-FEB-2002
L0921F05-3 NTA Mouse Newborn Kidney cDNA Library (Long) Mus
musculus cDNA clone L0921F05 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 412)
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
and Ko, M.S.H.
Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library
Unpublished (2001)
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0921 row: F column: 05

TITLE
JOURNAL
COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1402230.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:254074
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 359.

FEATURES
source

1..401
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:419522"
/sex="unknown"
/tissue_type="embryo"
/dev_stages="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NBM13.5 14.5"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pRT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 4; Length 412;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGGCCCGCCCTTGAAT 22
||||| ||| ||||| |||||
Db 303 TCCTTCCCTCGCCCTTGAAT 283

RESULT 41
BM224038/c
LOCUS
DEFINITION
BM224038 413 bp mRNA linear EST 07-JUN-2003
K0145A05-3 NTA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-)
cDNA Library (Long) Mus musculus cDNA clone NIA:K0145A05
IMAGE:30043684 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 413)
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
21429098
11544199
On Dec 14, 2001 this sequence version replaced gi:17785157.
Other ESTs: K0145A05-5N
Contact: Dawood B. Dudekula

Construction of long-transcript enriched cDNA libraries from

LOCUS	BY392061	420 bp	mRNA	linear	EST 12-DEC-2002
DEFINITION	BY392061	RIKEN full-length enriched, 13 days pregnant adult female placenta Mus musculus cDNA clone I530009P18			
ACCESSION	BY392061				
VERSION	BY392061.1	GI:26621549			
KEYWORDS	EST.				

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 420)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, F., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers
1. .420
/organism="Mus musculus"
/mol_type="mRNA"

Query Match 80.9%; Score 17.8; DB 5; Length 420;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
ORIGIN
QY 2 TCCTTGCCCCGCCCTTCAAT 22
||||| ||| ||||| |||||
DB 128 TCCTTGCCCCGCCCTTGAAT 148
RESULT 44
BM195973/c
LOCUS BM195973.2 GI:31522417
DEFINITION Mus musculus CDNA clone NIA:C0320E07 IMAGE:30006774 3', mRNA
ACCESSION BM195973
VERSION BM195973.2 GI:31522417
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
Genome Res. 11 (9), 1553-1558 (2001)
JOURNAL 21429098
MEDLINE 11544199
COMMENT On Dec 14, 2001 this sequence version replaced gi:17747648.
Other ESTs: C0320E07-5N
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0320 row: E column: 07
Seq primer: -21M13 Forward
High quality sequence stop: 422
POLYA=Yes.
Location/Qualifiers
1. .422
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129/Sv x 129/Sv-Cp"
/db_xref="niaEST:C0320E07-3"
/db_xref="taxon:10090"
/clones="NIA:C0320E07 IMAGE:30006774"
/tissue_type="Undifferentiated ES Cell"
/cell_line="R1 ES cells"
/lab_hosts="DH10B"
/clone_lib="NIA Mouse Undifferentiated ES Cell cDNA Library (Long)"
/note="Vector: pSPOT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001)). [EMBL: 11544199]. Total RNAs were obtained from Dr. Kenneth R. Boheler (National Institute on Aging, USA). ES cells were cultured without feeder cells in the presence of LIF and BRL-conditioned media.

Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen].
 5'-pGACTAGTCTAGATCGGAGCGCGCCCTTTT-3' from 14.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN

Query Match 80.9%; Score 17.8; DB 4; Length 422;
 Best Local Similarity 90.5%; Pred. No. 2.5e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
 ||||| ||||| ||||| ||||| |||||
 Db 293 TCCTTCCCTGCCCTTGAAT 273

RESULT 45

BM220960/c
 LOCUS BM220960 422 bp mRNA linear EST 07-JUN-2003
 DEFINITION C0947H10-3 NIA Mouse 12.5-dpc Male Genital Ridge/Mesonephros cDNA Library (Long) Mus musculus cDNA clone NIA:C0947H10 IMAGE:30039165 3', mRNA sequence.

ACCESSION

BM220960

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Piao, Y., Ko, N.T., Lin, M.K. and Ko, M.S.H.

Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

Genome Res. 11 (9), 1553-1558 (2001)

21429098

11544199

On Dec 14, 2001 this sequence version replaced gi:17781199.

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 400, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: C0947 row: H column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 422

POLYA=Yes

Location/Qualifiers

1..422

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="NIA:C0947H10 IMAGE:30039165"

/tissue_type="Male genital ridge/mesonephros"

/dev_stage="12.5-dpc"

/lab_host="DH10B"

/clone_lib="NIA Mouse 12.5-dpc Male Genital

Ridge/Mesonephros cDNA Library (Long)"

/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:

NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]:
 5'-pGACTAGTCTAGATCGGAGCGCGCCCTTTT-3' from 1.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN

Query Match 80.9%; Score 17.8; DB 4; Length 422;
 Best Local Similarity 90.5%; Pred. No. 2.5e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
 ||||| ||||| ||||| ||||| |||||
 Db 293 TCCTTCCCTGCCCTTGAAT 273

RESULT 46

BB731596

LOCUS BB731596 429 bp mRNA linear EST 12-OCT-2001

DEFINITION musculus cDNA clone E970012P06 3', mRNA sequence.

BB731596

ACCESSION BB731596.1 GI:16114871

VERSION

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,

Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A.,

Muramatsu, M., and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-rsg@sc.riken.jp, URL: <http://genome.sgc.riken.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
 Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.

FEATURES

source

1. 429
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E970012F06"
 /tissue_type="whole body"
 /dev_stage="12 days embryo"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 429;
 Best Local Similarity 90.5%; Pred. No. 2.5e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22

|||||
 143 TCCTTCCCTGCCCTTGAAT 163

RESULT 47

W89731

LOCUS

DEFINITION W89731 437 bp mRNA linear EST 12-SEP-1996
 m66606.ri Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
 clone IMAGE:419290 5' similar to TR:G1203895 G1203895 ISOPEPTIDASE
 T. [1] ;, mRNA sequence.

ACCESSION W89731

VERSION W89731.1

KEYWORDS GI:1542447

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 437)
 Maeda, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Fan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1405056.

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:253842

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 429.

FEATURES

source

1. 437
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"

/db_xref="taxon:10090"
 /clone="IMAGE:419290"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer 15'
 TGTACCAATCTGAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTTTTTT
 T 3', on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 p7T3 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Patima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 7; Length 437;
 Best Local Similarity 90.5%; Pred. No. 2.5e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22

|||||
 334 TCCTTCCCTGCCCTTGAAT 354

RESULT 48

W89731

LOCUS

DEFINITION W89731 442 bp mRNA linear EST 12-DEC-2002
 BY366980 RIKEN full-length enriched, 12 days embryo whole body Mus
 musculus cDNA clone E970026D15 3', mRNA sequence.

ACCESSION BY366980

VERSION BY366980.1

KEYWORDS GI:26596468

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 442)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaudo, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
 Clouthier, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, K., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.

REFERENCE

AUTHORS

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

22354683
12466851
MEDLINE
PUBMED
COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source

1. .442
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 12 days embryo
whole body"

ORIGIN

Query Match 80.9%; Score 17.8; DB 5; Length 442;
Best Local Similarity 90.5%; Pred. No. 2.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
|||||
Db 156 TCCTTGCCCGCCCTTGAAT 176
|||||

RESULT 49
BY474650

LOCUS BY474650 RIKEN full-length enriched, melanocyte Mus musculus cDNA
DEFINITION clone G270150P18 3', mRNA sequence.

ACCESSION BY474650

VERSION BY474650.1 GI:26809029

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE 1 (bases 1 to 452)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brucic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Meglott, D.R.,
Muta, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Numata, K., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Cells were provided by Drs. William J Pavan, Stacie Lofus, and
Denise Larson (Division of Intramural Research Genetic Disease
Research Branch National Human Genome Research Institute, National
Institutes of Health (NIH) Building: 49, Room 4A82 49 Convent Drive
MSC 4472 Bethesda, Maryland U.S.A) whose assistance we gratefully
acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
Location/Qualifiers
1. .452
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270150P18"

FEATURES
source

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:14:29 ; Search time 66 Seconds
(without alignments)
545.426 Million cell updates/sec

Title: US-10-010-476-14

Perfect score: 22

Sequence: 1 ATCCTGGCCCCCCCCCTTGAAAT 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813859 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUTS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	83.6	12163	4	US-09-949-016-16030
2	18.4	83.6	12164	4	US-09-949-016-12202
3	17.4	79.1	7517	4	US-09-949-016-15603
4	16.8	76.4	601	4	US-09-949-016-25546
5	16.8	76.4	601	4	US-09-949-016-30367
6	16.8	76.4	601	4	US-09-949-016-69656
7	16.8	76.4	601	4	US-09-949-016-77051
8	16.8	76.4	7080	4	US-09-589-034-12
9	16.8	76.4	11089	4	US-09-949-016-12381
10	16.8	76.4	11113	4	US-09-949-016-13791
11	16.8	76.4	20565	4	US-09-949-016-13125
12	16.8	76.4	20566	4	US-09-949-016-13990
13	16.8	76.4	50341	1	US-08-247-901C-1
14	16.8	76.4	50341	2	US-09-075-90A-1
15	16.8	76.4	52297	3	US-08-426-436-1
16	16.8	76.4	52297	3	US-08-705-557-1
17	16.4	74.5	4739	3	US-08-677-970-1
18	16.4	74.5	4403765	3	US-09-103-840A-2
19	16.4	74.5	4411529	3	US-09-103-840A-1
20	16.2	73.6	502	4	US-09-621-976-2480
21	16.2	73.6	506	4	US-09-621-976-614
22	16.2	73.6	601	4	US-09-949-016-48755
23	16.2	73.6	846	4	US-09-902-540-7470
24	16.2	73.6	1155	4	US-09-491-577-15
25	16.2	73.6	1354	4	US-09-491-577-106
26	16.2	73.6	5787	4	US-09-902-540-722
27	16.2	73.6	6882	4	US-09-949-016-4205

c 101	15.4	70.0	18558	4	US-09-949-016-13688	Sequence 13688, A	174	15.2	69.1	238815	4	US-09-949-016-16274	Sequence 16274, A
c 102	15.4	70.0	26700	1	US-08-472-217-1	Sequence 1, Appli	175	15.2	69.1	312470	4	US-09-949-016-14043	Sequence 14043, A
c 103	15.4	70.0	26700	2	US-08-488-199-5	Sequence 5, Appli	176	15.2	69.1	336024	4	US-09-949-016-12373	Sequence 12373, A
c 104	15.4	70.0	26700	3	US-08-760-53A-1	Sequence 1, Appli	c 177	15.2	69.1	422592	4	US-09-949-016-14182	Sequence 14182, A
c 105	15.4	70.0	26700	4	US-09-336-757-1	Sequence 1, Appli	178	15.2	69.1	4403765	3	US-09-103-840A-2	Sequence 1, Appli
c 106	15.4	70.0	57559	4	US-09-949-016-113077	Sequence 13077, A	179	15.2	69.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 107	15.4	70.0	57560	4	US-09-949-016-12536	Sequence 12536, A	c 180	15	68.2	601	4	US-09-949-016-94941	Sequence 94941, A
c 108	15.4	70.0	86380	4	US-09-949-016-14837	Sequence 14837, A	181	15	68.2	601	4	US-09-949-016-177930	Sequence 177930, A
c 109	15.2	69.1	187	4	US-09-513-993C-27535	Sequence 27535, A	182	15	68.2	601	4	US-09-949-016-177931	Sequence 177931, A
c 110	15.2	69.1	189	4	US-09-489-038A-6085	Sequence 6085, Ap	183	15	68.2	601	4	US-09-949-016-177937	Sequence 177937, A
c 111	15.2	69.1	255	4	US-09-248-796A-9081	Sequence 9081, Ap	184	15	68.2	601	4	US-09-949-016-177938	Sequence 177938, A
c 112	15.2	69.1	261	4	US-09-513-993C-32493	Sequence 32493, A	185	15	68.2	7846	4	US-09-949-016-14510	Sequence 14510, A
c 113	15.2	69.1	324	4	US-09-489-039A-5825	Sequence 5825, Ap	c 186	15	68.2	12249	4	US-09-949-016-16839	Sequence 16839, A
c 114	15.2	69.1	468	4	US-09-621-976-18950	Sequence 18950, A	c 187	15	68.2	12249	4	US-09-949-016-16840	Sequence 16840, A
c 115	15.2	69.1	487	4	US-09-621-976-17600	Sequence 17600, A	c 188	15	68.2	162465	4	US-09-949-016-14264	Sequence 14264, A
c 116	15.2	69.1	601	4	US-09-949-016-33690	Sequence 33690, A	189	14.8	67.3	94	4	US-09-626-002-28	Sequence 28, Appl
c 117	15.2	69.1	601	4	US-09-949-016-33691	Sequence 33691, A	190	14.8	67.3	95	4	US-09-626-002-27	Sequence 27, Appl
c 118	15.2	69.1	601	4	US-09-949-016-33697	Sequence 33697, A	191	14.8	67.3	113	4	US-09-626-002-26	Sequence 26, Appl
c 119	15.2	69.1	601	4	US-09-949-016-64697	Sequence 64697, A	192	14.8	67.3	137	4	US-09-626-002-25	Sequence 25, Appl
c 120	15.2	69.1	601	4	US-09-949-016-64698	Sequence 64698, A	193	14.8	67.3	153	4	US-09-626-002-24	Sequence 24, Appl
c 121	15.2	69.1	601	4	US-09-949-016-85839	Sequence 85839, A	c 194	14.8	67.3	153	4	US-09-626-002-30	Sequence 30, Appl
c 122	15.2	69.1	601	4	US-09-949-016-177932	Sequence 177932, A	c 195	14.8	67.3	192	4	US-09-248-796A-13990	Sequence 13990, A
c 123	15.2	69.1	712	1	US-08-340-539A-7	Sequence 7, Appli	c 196	14.8	67.3	209	4	US-09-626-002-31	Sequence 31, Appl
c 124	15.2	69.1	712	2	US-08-461-592B-7	Sequence 7, Appli	c 197	14.8	67.3	276	4	US-09-248-796A-13991	Sequence 13991, A
c 125	15.2	69.1	756	3	US-08-413-974-3	Sequence 3, Appli	c 198	14.8	67.3	296	4	US-09-626-002-23	Sequence 23, Appl
c 126	15.2	69.1	756	3	US-08-434-418-3	Sequence 3, Appli	c 199	14.8	67.3	296	4	US-09-626-002-29	Sequence 29, Appl
c 127	15.2	69.1	756	3	US-08-433-288-3	Sequence 3, Appli	c 200	14.8	67.3	328	4	US-09-621-976-13643	Sequence 13643, A
c 128	15.2	69.1	756	3	US-08-174-739A-3	Sequence 3, Appli	c 201	14.8	67.3	333	4	US-09-252-991A-6249	Sequence 6249, Ap
c 129	15.2	69.1	756	3	US-08-434-256-3	Sequence 3, Appli	c 202	14.8	67.3	396	4	US-09-513-999C-12076	Sequence 12076, A
c 130	15.2	69.1	810	3	US-08-413-974-5	Sequence 5, Appli	c 203	14.8	67.3	418	4	US-09-513-999C-3338	Sequence 3338, Ap
c 131	15.2	69.1	810	3	US-08-434-418-5	Sequence 5, Appli	c 204	14.8	67.3	445	4	US-09-626-002-22	Sequence 22, Appl
c 132	15.2	69.1	810	3	US-08-433-288-5	Sequence 5, Appli	c 205	14.8	67.3	502	4	US-09-626-002-20	Sequence 20, Appl
c 133	15.2	69.1	810	3	US-08-174-739A-5	Sequence 5, Appli	c 206	14.8	67.3	510	4	US-09-626-002-3	Sequence 3, Appli
c 134	15.2	69.1	810	3	US-08-434-256-5	Sequence 5, Appli	c 207	14.8	67.3	528	4	US-09-702-705-382	Sequence 382, App
c 135	15.2	69.1	948	4	US-09-902-540-3676	Sequence 3676, Ap	c 208	14.8	67.3	528	4	US-09-736-457-382	Sequence 382, App
c 136	15.2	69.1	1123	1	US-07-971-096-3	Sequence 3, Appli	c 209	14.8	67.3	528	4	US-09-614-124B-382	Sequence 382, App
c 137	15.2	69.1	1123	1	US-08-175-096-3	Sequence 3, Appli	c 210	14.8	67.3	528	4	US-09-671-325-382	Sequence 382, App
c 138	15.2	69.1	1452	4	US-09-614-221A-117	Sequence 317, App	c 211	14.8	67.3	528	4	US-09-589-184-382	Sequence 382, App
c 139	15.2	69.1	1509	4	US-09-252-991A-1602	Sequence 1602, Ap	c 212	14.8	67.3	528	4	US-09-658-824-382	Sequence 382, App
c 140	15.2	69.1	1572	4	US-09-252-991A-1420	Sequence 1420, Ap	c 213	14.8	67.3	547	4	US-09-573-080A-112	Sequence 112, App
c 141	15.2	69.1	1656	4	US-09-902-540-6531	Sequence 6531, Ap	c 214	14.8	67.3	601	4	US-09-949-016-21278	Sequence 21278, A
c 142	15.2	69.1	2181	4	US-09-252-991A-1700	Sequence 1700, Ap	c 215	14.8	67.3	601	4	US-09-949-016-23818	Sequence 23818, A
c 143	15.2	69.1	3822	4	US-09-949-016-4694	Sequence 4694, Ap	c 216	14.8	67.3	601	4	US-09-949-016-23868	Sequence 23868, A
c 144	15.2	69.1	3822	4	US-09-949-016-4695	Sequence 4695, Ap	c 217	14.8	67.3	601	4	US-09-949-016-23869	Sequence 23869, A
c 145	15.2	69.1	10182	4	US-09-902-540-1049	Sequence 1049, Ap	c 218	14.8	67.3	601	4	US-09-949-016-23870	Sequence 23870, A
c 146	15.2	69.1	12249	4	US-09-949-016-16839	Sequence 16839, A	c 219	14.8	67.3	601	4	US-09-949-016-32393	Sequence 32393, A
c 147	15.2	69.1	12249	4	US-09-949-016-16840	Sequence 16840, A	c 220	14.8	67.3	601	4	US-09-949-016-52216	Sequence 52216, A
c 148	15.2	69.1	12708	4	US-09-949-016-13216	Sequence 13216, A	c 221	14.8	67.3	601	4	US-09-949-016-55166	Sequence 55166, A
c 149	15.2	69.1	12928	4	US-09-949-016-13272	Sequence 12772, A	c 222	14.8	67.3	601	4	US-09-949-016-55167	Sequence 55167, A
c 150	15.2	69.1	12929	4	US-09-949-016-15702	Sequence 16702, A	c 223	14.8	67.3	601	4	US-09-949-016-55168	Sequence 55168, A
c 151	15.2	69.1	13634	4	US-09-949-016-13766	Sequence 16376, A	c 224	14.8	67.3	601	4	US-09-949-016-88482	Sequence 88482, A
c 152	15.2	69.1	16324	4	US-09-902-540-11778	Sequence 11778, Ap	c 225	14.8	67.3	601	4	US-09-949-016-128611	Sequence 128611, A
c 153	15.2	69.1	25621	4	US-09-902-540-1253	Sequence 1253, Ap	c 226	14.8	67.3	601	4	US-09-949-016-128631	Sequence 128631, A
c 154	15.2	69.1	29430	4	US-09-544-398B-7	Sequence 7, Appli	c 227	14.8	67.3	601	4	US-09-949-016-128642	Sequence 128642, A
c 155	15.2	69.1	29430	4	US-09-543-771B-7	Sequence 7, Appli	c 228	14.8	67.3	601	4	US-09-949-016-128653	Sequence 128653, A
c 156	15.2	69.1	33125	4	US-09-949-016-16436	Sequence 16436, A	c 229	14.8	67.3	601	4	US-09-949-016-183138	Sequence 183138, A
c 157	15.2	69.1	33125	4	US-09-949-016-16437	Sequence 16437, A	c 230	14.8	67.3	601	4	US-09-949-016-183139	Sequence 183139, A
c 158	15.2	69.1	48536	4	US-09-949-016-11867	Sequence 11867, A	c 231	14.8	67.3	606	3	US-09-328-111-386	Sequence 386, App
c 159	15.2	69.1	48536	4	US-09-949-016-17167	Sequence 17167, A	c 232	14.8	67.3	658	4	US-09-626-002-19	Sequence 19, Appl
c 160	15.2	69.1	49389	4	US-09-949-016-13547	Sequence 12547, A	c 233	14.8	67.3	741	4	US-09-621-976-3446	Sequence 3446, Ap
c 161	15.2	69.1	49389	4	US-09-949-016-13630	Sequence 13630, A	c 234	14.8	67.3	759	4	US-09-270-767-11732	Sequence 11732, A
c 162	15.2	69.1	51049	4	US-09-949-016-15571	Sequence 15571, A	c 235	14.8	67.3	765	4	US-09-252-991A-6050	Sequence 6050, Ap
c 163	15.2	69.1	54246	4	US-09-949-016-15206	Sequence 16206, A	c 236	14.8	67.3	1059	4	US-09-902-540-2450	Sequence 2450, Ap
c 164	15.2	69.1	74962	4	US-09-685-853A-3	Sequence 3, Appli	c 237	14.8	67.3	1182	1	US-07-640-476-4	Sequence 4, Appli
c 165	15.2	69.1	86380	4	US-09-949-016-15344	Sequence 15344, A	c 238	14.8	67.3	1185	6	5290690-1	Patent No. 5290690
c 166	15.2	69.1	87350	3	US-08-781-891-79	Sequence 79, Appl	c 239	14.8	67.3	1185	6	US-09-902-540-6915	Sequence 6915, Ap
c 167	15.2	69.1	87350	4	US-09-618-166-79	Sequence 79, Appl	c 240	14.8	67.3	1215	4	US-09-858-525B-7	Sequence 7, Appli
c 168	15.2	69.1	87543	3	US-09-791-211-3	Sequence 3, Appli	c 241	14.8	67.3	1239	4	US-09-858-525B-7	Sequence 7, Appli
c 169	15.2	69.1	93778	4	US-09-949-016-15096	Sequence 15096, A	c 242	14.8	67.3	1241	1	US-08-471-033-39	Sequence 39, Appl
c 170	15.2	69.1	139936	4	US-09-949-016-11782	Sequence 11782, A	c 243	14.8	67.3	1241	1	US-08-471-033-42	Sequence 42, Appl
c 171	15.2	69.1	139932	4	US-09-949-016-13280	Sequence 13280, A	c 244	14.8	67.3	1241	2	US-08-471-044-39	Sequence 39, Appl
c 172	15.2	69.1	140844	4	US-09-949-016-14199	Sequence 14199, A	c 245	14.8	67.3	1241	2	US-08-471-044-42	Sequence 42, Appl
c 173	15.2	69.1	153866	4	US-09-949-016-16919	Sequence 16919, A	c 246	14.8	67.3	1241	2	US-08-463-483A-39	Sequence 39, Appl

C 247	14.8	67.3	1241	2	US-08-463-483A-42	Sequence 42, Appl	C 320	14.8	67.3	7804	4	US-09-949-016-1241	Sequence 14241, A
C 248	14.8	67.3	1241	2	US-08-471-046A-39	Sequence 39, Appl	C 321	14.8	67.3	8436	4	US-09-949-016-13963	Sequence 13963, A
C 249	14.8	67.3	1241	2	US-08-471-046A-42	Sequence 42, Appl	C 322	14.8	67.3	10434	4	US-09-949-016-15388	Sequence 15388, A
C 250	14.8	67.3	1241	2	US-08-470-566B-39	Sequence 39, Appl	C 323	14.8	67.3	10434	4	US-09-949-016-15389	Sequence 15389, A
C 251	14.8	67.3	1241	2	US-08-470-566B-42	Sequence 42, Appl	C 324	14.8	67.3	12567	4	US-09-949-016-16843	Sequence 16843, A
C 252	14.8	67.3	1241	2	US-08-469-334-39	Sequence 39, Appl	C 325	14.8	67.3	14438	4	US-09-949-016-12300	Sequence 12300, A
C 253	14.8	67.3	1241	2	US-08-469-334-42	Sequence 42, Appl	C 326	14.8	67.3	14484	4	US-09-902-540-1108	Sequence 1108, Ap
C 254	14.8	67.3	1241	3	US-09-300-529-39	Sequence 39, Appl	C 327	14.8	67.3	14515	4	US-09-948-016-11915	Sequence 11915, A
C 255	14.8	67.3	1241	3	US-09-300-529-42	Sequence 42, Appl	C 328	14.8	67.3	14621	4	US-09-949-016-15386	Sequence 15386, A
C 256	14.8	67.3	1241	3	US-09-252-991A-5147	Sequence 5147, Ap	C 329	14.8	67.3	14621	4	US-09-949-016-15387	Sequence 15387, A
C 257	14.8	67.3	1358	1	US-08-471-033-45	Sequence 45, Appl	C 330	14.8	67.3	14989	4	US-09-949-016-17089	Sequence 17089, A
C 258	14.8	67.3	1358	2	US-08-471-044-45	Sequence 45, Appl	C 331	14.8	67.3	15116	4	US-09-949-016-12354	Sequence 12354, A
C 259	14.8	67.3	1358	2	US-08-463-483A-45	Sequence 45, Appl	C 332	14.8	67.3	15117	4	US-09-949-016-16260	Sequence 16260, A
C 260	14.8	67.3	1358	2	US-08-471-046A-45	Sequence 45, Appl	C 333	14.8	67.3	15273	4	US-09-949-016-12356	Sequence 12356, A
C 261	14.8	67.3	1358	2	US-08-470-566B-45	Sequence 45, Appl	C 334	14.8	67.3	15273	4	US-09-949-016-13341	Sequence 13341, A
C 262	14.8	67.3	1358	2	US-08-469-334-45	Sequence 45, Appl	C 335	14.8	67.3	15412	4	US-09-949-016-15869	Sequence 15869, A
C 263	14.8	67.3	1358	2	US-09-300-529-45	Sequence 45, Appl	C 336	14.8	67.3	15461	4	US-09-949-016-17604	Sequence 17604, A
C 264	14.8	67.3	1362	3	US-08-979-608A-12	Sequence 12, Appl	C 337	14.8	67.3	17353	4	US-09-949-016-13437	Sequence 13437, A
C 265	14.8	67.3	1362	4	US-09-517-849-12	Sequence 12, Appl	C 338	14.8	67.3	17503	4	US-09-902-540-1114	Sequence 1114, Ap
C 266	14.8	67.3	1362	4	US-09-616-289-12	Sequence 12, Appl	C 339	14.8	67.3	17503	4	US-09-948-016-13250	Sequence 13250, A
C 267	14.8	67.3	1389	1	US-08-471-033-27	Sequence 27, Appl	C 340	14.8	67.3	17503	4	US-09-949-016-12690	Sequence 12690, A
C 268	14.8	67.3	1389	2	US-08-471-044-27	Sequence 27, Appl	C 341	14.8	67.3	17503	4	US-09-949-016-16195	Sequence 16195, A
C 269	14.8	67.3	1389	2	US-08-463-483A-27	Sequence 27, Appl	C 342	14.8	67.3	17503	4	US-09-949-016-12179	Sequence 12179, A
C 270	14.8	67.3	1389	2	US-08-471-046A-27	Sequence 27, Appl	C 343	14.8	67.3	17503	4	US-09-949-016-14472	Sequence 14472, A
C 271	14.8	67.3	1389	2	US-08-470-566B-27	Sequence 27, Appl	C 344	14.8	67.3	17503	4	US-09-949-016-12494	Sequence 12494, A
C 272	14.8	67.3	1389	2	US-08-469-334-27	Sequence 27, Appl	C 345	14.8	67.3	17503	4	US-09-949-016-14726	Sequence 14726, A
C 273	14.8	67.3	1389	2	US-09-300-529-27	Sequence 27, Appl	C 346	14.8	67.3	17503	4	US-09-949-016-16318	Sequence 16318, A
C 274	14.8	67.3	1399	1	US-08-471-033-24	Sequence 24, Appl	C 347	14.8	67.3	17503	4	US-09-949-016-16988	Sequence 16988, A
C 275	14.8	67.3	1399	2	US-08-471-044-24	Sequence 24, Appl	C 348	14.8	67.3	17503	4	US-09-949-016-16956	Sequence 16956, A
C 276	14.8	67.3	1399	2	US-08-463-483A-24	Sequence 24, Appl	C 349	14.8	67.3	17503	4	US-09-949-016-16664	Sequence 16664, A
C 277	14.8	67.3	1399	2	US-08-471-046A-24	Sequence 24, Appl	C 350	14.8	67.3	17503	4	US-09-676-610B-24	Sequence 24, Appl
C 278	14.8	67.3	1399	2	US-08-470-566B-24	Sequence 24, Appl	C 351	14.8	67.3	17503	4	US-09-949-016-16513	Sequence 16513, A
C 279	14.8	67.3	1399	2	US-08-469-334-24	Sequence 24, Appl	C 352	14.8	67.3	17503	4	US-09-877-177A-10	Sequence 10, Appl
C 280	14.8	67.3	1399	2	US-09-300-529-24	Sequence 24, Appl	C 353	14.8	67.3	17503	4	US-09-949-016-14516	Sequence 14516, A
C 281	14.8	67.3	1422	3	US-08-979-608A-13	Sequence 13, Appl	C 354	14.8	67.3	17503	4	US-09-949-016-14517	Sequence 14517, A
C 282	14.8	67.3	1422	4	US-09-517-849-13	Sequence 13, Appl	C 355	14.8	67.3	17503	4	US-09-949-016-14518	Sequence 14518, A
C 283	14.8	67.3	1422	4	US-09-517-849-13	Sequence 13, Appl	C 356	14.8	67.3	17503	4	US-09-949-016-14519	Sequence 14519, A
C 284	14.8	67.3	1422	4	US-09-616-289-13	Sequence 13, Appl	C 357	14.8	67.3	17503	4	US-09-949-016-17226	Sequence 17226, A
C 285	14.8	67.3	1434	2	US-09-055-097-2	Sequence 2, Appl	C 358	14.8	67.3	17503	4	US-09-949-016-17227	Sequence 17227, A
C 286	14.8	67.3	1434	2	US-09-373-902-2	Sequence 2, Appl	C 359	14.8	67.3	17503	4	US-09-949-016-17228	Sequence 17228, A
C 287	14.8	67.3	1467	4	US-09-252-991A-5070	Sequence 5070, Ap	C 360	14.8	67.3	17503	4	US-09-949-016-17229	Sequence 17229, A
C 288	14.8	67.3	1522	4	US-09-949-016-1595	Sequence 1595, Ap	C 361	14.8	67.3	17503	4	US-09-949-016-14193	Sequence 14193, A
C 289	14.8	67.3	1566	6	5290690-3	Patent No. 5290690	C 362	14.8	67.3	17503	4	US-09-949-016-12900	Sequence 12900, A
C 290	14.8	67.3	1566	6	5290690-3	Patent No. 5290690	C 363	14.8	67.3	17503	4	US-09-949-016-12412	Sequence 12412, A
C 291	14.8	67.3	1617	3	US-08-979-608A-11	Sequence 11, Appl	C 364	14.8	67.3	17503	4	US-09-949-016-12412	Sequence 12412, A
C 292	14.8	67.3	1617	3	US-09-517-849-11	Sequence 11, Appl	C 365	14.8	67.3	17503	4	US-09-702-705-1597	Sequence 1597, Ap
C 293	14.8	67.3	1617	4	US-09-616-289-11	Sequence 11, Appl	C 366	14.8	67.3	17503	4	US-09-736-457-1597	Sequence 1597, Ap
C 294	14.8	67.3	1640	4	US-09-949-016-5696	Sequence 5696, Ap	C 367	14.8	67.3	17503	4	US-09-614-124B-1597	Sequence 1597, Ap
C 295	14.8	67.3	1644	4	US-09-902-540-371	Sequence 371, App	C 368	14.8	67.3	17503	4	US-09-671-325-1597	Sequence 1597, Ap
C 296	14.8	67.3	1689	4	US-09-902-540-2528	Sequence 2528, Ap	C 369	14.8	67.3	17503	4	US-09-658-824-1597	Sequence 1597, Ap
C 297	14.8	67.3	2122	4	US-09-949-016-614	Sequence 614, App	C 370	14.8	67.3	17503	4	US-09-513-999C-21816	Sequence 21816, A
C 298	14.8	67.3	2122	4	US-09-949-016-1599	Sequence 1599, Ap	C 371	14.8	67.3	17503	4	US-08-507-431-40	Sequence 40, Appl
C 299	14.8	67.3	2309	3	US-09-091-725-9	Sequence 9, Appl	C 372	14.8	67.3	17503	4	US-09-116-622-40	Sequence 40, Appl
C 300	14.8	67.3	2561	4	US-09-616-289-48	Sequence 48, Appl	C 373	14.8	67.3	17503	4	US-09-116-622-40	Sequence 40, Appl
C 301	14.8	67.3	3016	4	US-09-023-653-259	Sequence 259, App	C 374	14.8	67.3	17503	4	US-09-219-277-40	Sequence 40, Appl
C 302	14.8	67.3	4031	1	US-08-471-033-49	Sequence 49, Appl	C 375	14.8	67.3	17503	4	US-09-583-110-325	Sequence 325, App
C 303	14.8	67.3	4031	2	US-08-471-044-49	Sequence 49, Appl	C 376	14.8	67.3	17503	4	US-09-313-294A-2062	Sequence 2062, Ap
C 304	14.8	67.3	4031	2	US-08-463-483A-49	Sequence 49, Appl	C 377	14.8	67.3	17503	4	US-09-513-999C-11831	Sequence 11831, A
C 305	14.8	67.3	4031	2	US-08-471-046A-49	Sequence 49, Appl	C 378	14.8	67.3	17503	4	US-09-270-767-27637	Sequence 27637, A
C 306	14.8	67.3	4031	2	US-08-470-566B-49	Sequence 49, Appl	C 379	14.8	67.3	17503	4	US-09-270-767-27637	Sequence 27637, A
C 307	14.8	67.3	4031	2	US-08-469-334-49	Sequence 49, Appl	C 380	14.8	67.3	17503	4	US-09-270-767-27637	Sequence 27637, A
C 308	14.8	67.3	4031	3	US-09-300-529-49	Sequence 49, Appl	C 381	14.8	67.3	17503	4	US-09-513-999C-3966	Sequence 3966, Ap
C 309	14.8	67.3	4812	4	US-09-949-016-17058	Sequence 17058, A	C 382	14.8	67.3	17503	4	US-09-621-976-1160	Sequence 1160, Ap
C 310	14.8	67.3	4829	4	US-09-902-540-592	Sequence 592, App	C 383	14.8	67.3	17503	4	US-09-895-652A-18	Sequence 18, Appl
C 311	14.8	67.3	5217	4	US-09-573-080A-201	Sequence 201, App	C 384	14.8	67.3	17503	4	US-09-949-016-43091	Sequence 43091, A
C 312	14.8	67.3	5881	4	US-09-949-016-12049	Sequence 12049, A	C 385	14.8	67.3	17503	4	US-09-949-016-55246	Sequence 55246, A
C 313	14.8	67.3	5881	4	US-09-902-540-7351	Sequence 7351, Ap	C 386	14.8	67.3	17503	4	US-09-949-016-60305	Sequence 60305, A
C 314	14.8	67.3	5881	4	US-09-949-016-15589	Sequence 15589, A	C 387	14.8	67.3	17503	4	US-09-949-016-68942	Sequence 68942, A
C 315	14.8	67.3	5881	4	US-09-949-016-17438	Sequence 17438, A	C 388	14.8	67.3	17503	4	US-09-949-016-68943	Sequence 68943, A
C 316	14.8	67.3	5971	4	US-09-373-275-26	Sequence 26, Appl	C 389	14.8	67.3	17503	4	US-09-949-016-68944	Sequence 68944, A
C 317	14.8	67.3	6402	4	US-09-902-540-702	Sequence 702, App	C 390	14.8	67.3	17503	4	US-09-949-016-68973	Sequence 68973, A
C 318	14.8	67.3	7261	4	US-09-949-016-14780	Sequence 14780, A	C 391	14.8	67.3	17503	4	US-09-949-016-79935	Sequence 79935, A
C 319	14.8	67.3	7803	4	US-09-949-016-12491	Sequence 12491, A	C 392	14.8	67.3	17503	4	US-09-949-016-79936	Sequence 79936, A

C 393	14.6	65.4	601	4	US-09-949-016-81126	Sequence 81126, A	C 466	14.6	66.4	3243	4	US-08-468-581A-6	Sequence 6, Appli
C 394	14.6	66.4	601	4	US-09-949-016-81127	Sequence 81127, A	C 467	14.6	66.4	3478	4	US-09-949-016-5257	Sequence 5257, Ap
C 395	14.6	66.4	601	4	US-09-949-016-91767	Sequence 91767, A	C 468	14.6	66.4	3493	4	US-09-949-016-780	Sequence 780, App
C 396	14.6	66.4	601	4	US-09-949-016-91768	Sequence 91768, A	C 469	14.6	66.4	6273	3	US-08-961-527-21	Sequence 21, Appli
C 397	14.6	66.4	601	4	US-09-949-016-120050	Sequence 120050, A	C 470	14.6	66.4	7353	4	US-09-949-016-14895	Sequence 14895, A
C 398	14.6	66.4	601	4	US-09-949-016-120100	Sequence 120100, A	C 471	14.6	66.4	8461	4	US-09-949-016-13428	Sequence 13428, A
C 399	14.6	66.4	601	4	US-09-949-016-120150	Sequence 120150, A	C 472	14.6	66.4	9301	3	US-09-449-218D-18	Sequence 18, Appli
C 400	14.6	66.4	601	4	US-09-949-016-127492	Sequence 127492, A	C 473	14.6	66.4	9301	4	US-09-668-529A-18	Sequence 18, Appli
C 401	14.6	66.4	601	4	US-09-949-016-128441	Sequence 128441, A	C 474	14.6	66.4	9301	4	US-09-668-037A-18	Sequence 18, Appli
C 402	14.6	66.4	601	4	US-09-949-016-136242	Sequence 136242, A	C 475	14.6	66.4	9301	4	US-09-668-021-18	Sequence 18, Appli
C 403	14.6	66.4	601	4	US-09-949-016-142664	Sequence 142664, A	C 476	14.6	66.4	11105	4	US-09-949-016-13218	Sequence 13218, A
C 404	14.6	66.4	601	4	US-09-949-016-175921	Sequence 175921, A	C 477	14.6	66.4	12412	1	US-08-390-878-18	Sequence 18, Appli
C 405	14.6	66.4	601	4	US-09-949-016-175901	Sequence 175901, A	C 478	14.6	66.4	14185	4	US-09-949-016-14809	Sequence 14809, A
C 406	14.6	66.4	601	4	US-09-949-016-189169	Sequence 189169, A	C 479	14.6	66.4	17085	4	US-09-949-016-16507	Sequence 16507, A
C 407	14.6	66.4	601	4	US-09-949-016-192839	Sequence 192839, A	C 480	14.6	66.4	17138	3	US-09-813-819-3	Sequence 3, Appli
C 408	14.6	66.4	601	4	US-09-949-016-194708	Sequence 194708, A	C 481	14.6	66.4	17138	3	US-09-920-048-3	Sequence 3, Appli
C 409	14.6	66.4	601	4	US-09-949-016-194709	Sequence 194709, A	C 482	14.6	66.4	17138	4	US-10-014-501-3	Sequence 3, Appli
C 410	14.6	66.4	685	4	US-09-513-999C-13230	Sequence 13230, A	C 483	14.6	66.4	17644	4	US-09-949-016-16342	Sequence 16342, A
C 411	14.6	66.4	729	4	US-09-583-110-324	Sequence 324, App	C 484	14.6	66.4	17645	4	US-09-949-016-11805	Sequence 11805, A
C 412	14.6	66.4	738	4	US-09-107-433-2224	Sequence 2224, App	C 485	14.6	66.4	17645	4	US-09-949-016-14188	Sequence 14188, A
C 413	14.6	66.4	783	4	US-09-253-991A-16173	Sequence 16173, A	C 486	14.6	66.4	19009	4	US-09-949-016-12522	Sequence 12522, A
C 414	14.6	66.4	828	4	US-09-540-236-113	Sequence 113, App	C 487	14.6	66.4	19012	4	US-09-949-016-16999	Sequence 16999, A
C 415	14.6	66.4	905	4	US-09-270-767-11679	Sequence 11679, A	C 488	14.6	66.4	23928	4	US-09-949-016-12392	Sequence 12392, A
C 416	14.6	66.4	966	4	US-09-602-787A-129	Sequence 129, App	C 489	14.6	66.4	23928	4	US-09-949-016-16315	Sequence 16315, A
C 417	14.6	66.4	1208	4	US-09-602-787A-409	Sequence 409, App	C 490	14.6	66.4	25769	4	US-09-949-016-17154	Sequence 17154, A
C 418	14.6	66.4	1247	4	US-09-270-767-11477	Sequence 11477, A	C 491	14.6	66.4	26763	4	US-09-949-016-11984	Sequence 11984, A
C 419	14.6	66.4	1273	1	US-08-507-431-1	Sequence 1, Appli	C 492	14.6	66.4	26764	4	US-09-949-016-15566	Sequence 15566, A
C 420	14.6	66.4	1273	2	US-08-902-655A-1	Sequence 1, Appli	C 493	14.6	66.4	29272	4	US-09-902-540-1217	Sequence 1217, Ap
C 421	14.6	66.4	1273	3	US-09-116-622-1	Sequence 1, Appli	C 494	14.6	66.4	30324	4	US-09-949-016-16037	Sequence 16037, A
C 422	14.6	66.4	1273	3	US-09-219-277-1	Sequence 1, Appli	C 495	14.6	66.4	31147	4	US-09-596-002-25	Sequence 25, Appli
C 423	14.6	66.4	1273	3	US-09-599-661-1	Sequence 1, Appli	C 496	14.6	66.4	32030	4	US-09-949-016-13127	Sequence 13127, A
C 424	14.6	66.4	1280	4	US-09-602-787A-407	Sequence 407, App	C 497	14.6	66.4	34230	4	US-09-949-016-12052	Sequence 12052, A
C 425	14.6	66.4	1322	3	US-08-973-005A-1	Sequence 1, Appli	C 498	14.6	66.4	34276	4	US-09-949-016-12263	Sequence 12263, A
C 426	14.6	66.4	1347	4	US-09-016-434-1261	Sequence 1261, App	C 499	14.6	66.4	34278	4	US-09-949-016-16103	Sequence 16103, A
C 427	14.6	66.4	1397	2	US-08-239-276-1	Sequence 1, Appli	C 500	14.6	66.4	38969	4	US-09-949-016-14051	Sequence 14051, A
C 428	14.6	66.4	1397	2	US-08-468-579B-1	Sequence 1, Appli							
C 429	14.6	66.4	1397	3	US-08-468-577B-1	Sequence 1, Appli							
C 430	14.6	66.4	1397	4	US-08-468-581A-1	Sequence 1, Appli							
C 431	14.6	66.4	1528	4	US-09-270-767-11960	Sequence 11960, A							
C 432	14.6	66.4	1570	2	US-08-239-276-5	Sequence 5, Appli							
C 433	14.6	66.4	1570	2	US-08-468-579B-5	Sequence 5, Appli							
C 434	14.6	66.4	1570	3	US-08-468-577B-5	Sequence 5, Appli							
C 435	14.6	66.4	1570	4	US-08-468-581A-5	Sequence 5, Appli							
C 436	14.6	66.4	1607	3	US-09-853-768-13	Sequence 13, Appli							
C 437	14.6	66.4	1652	1	US-08-036-555B-148	Sequence 148, App							
C 438	14.6	66.4	1652	1	US-08-463-569-148	Sequence 148, App							
C 439	14.6	66.4	1652	1	US-08-249-322A-148	Sequence 148, App							
C 440	14.6	66.4	1652	1	US-08-469-526A-148	Sequence 148, App							
C 441	14.6	66.4	1652	1	US-08-734-591A-148	Sequence 148, App							
C 442	14.6	66.4	1652	2	US-08-469-660-148	Sequence 148, App							
C 443	14.6	66.4	1652	3	US-08-341-018-53	Sequence 53, Appli							
C 444	14.6	66.4	1652	3	US-08-470-335-148	Sequence 148, App							
C 445	14.6	66.4	1652	3	US-08-735-021-148	Sequence 148, App							
C 446	14.6	66.4	1652	3	US-08-734-664A-148	Sequence 148, App							
C 447	14.6	66.4	1652	4	US-08-470-339-148	Sequence 148, App							
C 448	14.6	66.4	1652	3	US-08-467-602-148	Sequence 148, App							
C 449	14.6	66.4	1652	4	US-08-411-295F-46	Sequence 46, Appli							
C 450	14.6	66.4	1652	5	PCT-US94-05083C-144	Sequence 144, App							
C 451	14.6	66.4	1652	5	PCT-US95-06846A-148	Sequence 148, App							
C 452	14.6	66.4	2306	6	5198359-3	Patent No. 5198359							
C 453	14.6	66.4	2306	6	5449756-3	Patent No. 5449756							
C 454	14.6	66.4	2306	6	5198359-3	Patent No. 5198359							
C 455	14.6	66.4	2306	6	5449756-3	Patent No. 5449756							
C 456	14.6	66.4	2329	1	US-08-455-559-9	Sequence 9, Appli							
C 457	14.6	66.4	2329	5	US-09-145-060-9	Sequence 9, Appli							
C 458	14.6	66.4	2329	5	PCT-US94-00657-9	Sequence 9, Appli							
C 459	14.6	66.4	2540	4	US-09-023-655-295	Sequence 295, App							
C 460	14.6	66.4	2704	4	US-09-220-132-27	Sequence 27, Appli							
C 461	14.6	66.4	2967	4	US-09-902-540-4397	Sequence 4397, App							
C 462	14.6	66.4	3120	3	US-09-221-017B-54	Sequence 54, Appli							
C 463	14.6	66.4	3243	2	US-08-239-276-6	Sequence 6, Appli							
C 464	14.6	66.4	3243	2	US-08-468-579B-6	Sequence 6, Appli							
C 465	14.6	66.4	3243	3	US-08-468-577B-6	Sequence 6, Appli							
													Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-16030

; Sequence 16030, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16030

; LENGTH: 12163

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-16030

Query Match 83.6%; Score 18.4; DB 4; Length 12163;

Best Local Similarity 95.0%; Pred. No. 46;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCCCCTTGAAT 22

Db 1499 CCGTGGCCCCCCTTGAAT 1518


```
RESULT 2
US-09-949-016-12202
; Sequence 12202, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12202
; LENGTH: 12164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12202

Query Match      83.6%; Score 18.4; DB 4; Length 12164;
Best Local Similarity 95.0%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CCTGTGCCCGCCCTTGAAT 22
DB      1499 CCGTGCCCGCCCTTGAAT 1518

RESULT 3
US-09-949-016-15603
; Sequence 15603, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15603
; LENGTH: 7517
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15603

Query Match      79.1%; Score 17.4; DB 4; Length 7517;
Best Local Similarity 94.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATCCTTGCCCGCCCTTGAAT 19
DB      6147 ATCCTTGCCCGCCCTTGAAT 6165

RESULT 4
US-09-949-016-25546
; Sequence 25546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25546
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-25546

Query Match      76.4%; Score 16.8; DB 4; Length 601;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 CCTTGCCCGCCCTTGAAT 22
DB      435 CCTTGCCCGCCCTTGAAT 454

RESULT 5
US-09-949-016-30367
; Sequence 30367, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30367
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30367

Query Match      76.4%; Score 16.8; DB 4; Length 601;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ATCCTTGCCCGCCCTTGA 20
DB      35 ATCCTTGTCGCCCGCCCTGGA 54

RESULT 6
US-09-949-016-69656
; Sequence 69656, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69656
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-69656

Query Match      76.4%; Score 16.8; DB 4; Length 601;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 CCTTGCCCGCCCTTGAAT 22
DB      435 CCTTGCCCGCCCTTGAAT 454
```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69656
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-69656

```

```
Query Match          76.4%; Score 16.8; DB 4; Length 601;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 3 CCTTGCCCGCCCTTGAAT 22
|||
Db 435 CCCTGCTCCGCCCTTGAAT 454

RESULT 7
US-09-949-016-77051
Sequence 77051, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 77051
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-77051

```
Query Match          76.4%; Score 16.8; DB 4; Length 601;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 ATCCTTGCCCGGCCCTTGA 20
|||
pb 35 ATCCTTGTCGCCGCCCTTGA 54
|||

```

RESULT 8
US-09-589-034-12
; Sequence 12, Application US/09589034
; Patent No. 6699983
; GENERAL INFORMATION:
; APPLICANT: Dedhar, Shoukat
; APPLICANT: Hannigan, Greg
; TITLE OF INVENTION: Integrin-linked Kinase and its Uses
; FILE REFERENCE: KINE-001CIP3
; CURRENT APPLICATION NUMBER: US/09/589,034
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 60/009,074
; PRIOR FILING DATE: 1995-12-21
; PRIOR APPLICATION NUMBER: 08/752,345

```

```

; PRIOR FILING DATE: 1996-11-19
; PRIOR APPLICATION NUMBER: 08/955,841
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 09/935,706
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: 09/390,425
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 7080
; TYPE: DNA

```

```
Query Match          76.4%; Score 16.8; DB 4; Length 7080;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 3 CCTTGCCCGCCCTTGAAT 22
Dp 3366 CCTGTCCGCCCTTGAAT 3385

```

RESULT 9
US-09-949-016-12381
; Sequence 12381, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12381
; LENGTH: 11089
; TYPE: DNA

```

```
; ORGANISM: Human
US-09-949-016-12381

Query Match          76.4%; Score 16.8; DB 4; Length 11089;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTGTGCGCCGCCCTTGAAT 22
DB 5375 CCTGTGCGCCGCCCTTGAAT 5394

RESULT 10
US-09-949-016-13791
; Sequence 13791, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13791
; LENGTH: 11113
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13791

Query Match          76.4%; Score 16.8; DB 4; Length 11113;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTGTGCGCCGCCCTTGAAT 22
DB 5394 CCTGTGCGCCGCCCTTGAAT 5413

RESULT 11
US-09-949-016-12125
; Sequence 12125, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12125
; LENGTH: 20565
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12125

Query Match          76.4%; Score 16.8; DB 4; Length 20565;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCGCCCTTGA 20
DB 3290 ATCCTTGTCGCCGCCCTTGA 3309

RESULT 12
US-09-949-016-13990
; Sequence 13990, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13990
; LENGTH: 20566
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13990

Query Match          76.4%; Score 16.8; DB 4; Length 20566;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCGCCCTTGA 20
DB 3290 ATCCTTGTCGCCGCCCTTGA 3309

RESULT 13
US-08-247-901C-1/c
; Sequence 1, Application US/08247901C
; Patent No. 5750384
; GENERAL INFORMATION:
; APPLICANT: Jacobs et al
; TITLE OF INVENTION: LS SHUTTLE PHASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,901C
; FILING DATE: May 23, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/057,531
; FILING DATE: April 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Elizabeth A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/273
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 50341
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: L5 shuttle phasmid sequence
DESCRIPTION: L5 shuttle phasmid sequence
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: L5 mycobacteriophage
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5750384e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1

Query Match 76.4%; Score 16.8; DB 1; Length 50341;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20
|||||
DB 7079 ATCCTTGCCCGCCCTTGA 7060

RESULT 14
US-09-075-904-1/c
Sequence 1, Application US/09075904
GENERAL INFORMATION:
APPLICANT: Jacobs, et al.
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS

SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,904
FILING DATE: May 11, 1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,901
FILING DATE: May 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Elizabeth A
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/475
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 50341
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: L5 shuttle phasmid sequence
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: L5 mycobacteriophage
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5994137e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1

Query Match 76.4%; Score 16.8; DB 2; Length 50341;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20
|||||
DB 7079 ATCCTTGCCCGCCCTTGA 7060

RESULT 15
US-09-426-436-1/c
Sequence 1, Application US/09426436
Patent No. 6225066
GENERAL INFORMATION:
APPLICANT: William R. Jacobs, Jr.

APPLICANT: Barry R. Bloom
APPLICANT: Graham F. Hatfull
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/426,436
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/705,557
FILING DATE:
APPLICATION NUMBER: US/08/057,531
FILING DATE:
APPLICATION NUMBER: 07/833,431
FILING DATE: February 7, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 96700/238
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: phage genome sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
INDIVIDUAL ISOLATE: L5
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL LINE: not applicable
ORGANELLE: not applicable
IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
TITLE: Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 395-405

DATE: 1993
US-09-426-436-1
Query Match 76.4%; Score 16.8; DB 3; Length 52297;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATCCTTGCCCGCCCTTGA 20
DB 6976 ATCCTTGCCCGCCCTTGA 6957
RESULT 16
US-08-705-557-1/c
Sequence 1, Application US/08705557
Patent No. 6300061
GENERAL INFORMATION:
APPLICANT: William R. Jacobs, Jr.
APPLICANT: Barry R. Bloom
APPLICANT: Graham F. Hatfull
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705,557
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/057,531
FILING DATE:
APPLICATION NUMBER: 07/833,431
FILING DATE: February 7, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 96700/238
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: phage genome sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
INDIVIDUAL ISOLATE: L5
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL TYPE: not applicable
CELL LINE: not applicable
ORGANELLE: not applicable

IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
TITLE: Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
TITLE: Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 395-405
DATE: 1993
US-08-705-557-1

Query Match 76.4%; Score 16.8; DB 3; Length 52297;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCTTGCCCCGCCCTTGA 20
|||||
DB 6976 ATCTTGCCCCGCCCTTGA 6957

RESULT 17
US-08-677-970-1/c
; Sequence 1, Application US/08677970
; Patent No. 644444
; GENERAL INFORMATION:
; APPLICANT: Anand, Naveen N.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: GENES ENCODING MYCOBACTERIAL PROTEINS
; TITLE OF INVENTION: ASSOCIATED WITH CELL BINDING AND CELL ENTRY AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1K7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/677,970
FILING DATE: 10-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-592 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4739 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

Query Match 74.5%; Score 16.4; DB 3; Length 4739;
Best Local Similarity 94.4%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCTTG 19
|||||
DB 282 TCCTTGCCCCGCCCTTG 265

RESULT 18
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 74.5%; Score 16.4; DB 3; Length 4403765;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCTTG 19
|||||
DB 196960 TCCTTGCCCCGCCCTTG 196943

RESULT 19
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 74.5%; Score 16.4; DB 3; Length 4411529;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCTTG 19
|||||
DB 196793 TCCTTGCCCCGCCCTTG 196776

```
RESULT 20
US-09-621-976-2480/c
; Sequence 2480, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2480
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124...489
US-09-621-976-2480

Query Match          73.6%; Score 16.2; DB 4; Length 502;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 401 TCTTGCTTCCCTTGAAT 381

RESULT 21
US-09-621-976-614/c
; Sequence 614, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 614
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 187...504
US-09-621-976-614

Query Match          73.6%; Score 16.2; DB 4; Length 506;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 464 TCTTGCTTCCCTTGAAT 444

RESULT 22
US-09-949-016-48755/c
; Sequence 48755, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48755
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48755

Query Match          73.6%; Score 16.2; DB 4; Length 601;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 26 TCCTTGCCCGCCCTTGAAT 6

RESULT 23
US-09-902-540-7470/c
; Sequence 7470, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7470
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7470

Query Match          73.6%; Score 16.2; DB 4; Length 846;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 21
Db 219 ATCCTTGCCCGCCCTTGA 199

RESULT 24
US-09-491-577-15/c
; Sequence 15, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
; APPLICANT: Yale University
; APPLICANT: Carlson, John R.
; APPLICANT: Kim, Hunhyong
; APPLICANT: Clyne, Peter J.
; APPLICANT: Marr, Coral G.
; TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
; FILE REFERENCE: 44574-5061-US
; CURRENT APPLICATION NUMBER: US/09/491,577
; CURRENT FILING DATE: 2000-01-25
; EARLIER APPLICATION NUMBER: US 60/117,132
; EARLIER FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 112
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1152)
; OTHER INFORMATION: DOR 33B3.3, a coding region on BDGP Clone No. 6610511
; OTHER INFORMATION: AC006240
US-09-491-577-15

Query Match      73.6%; Score 16.2; DB 4; Length 1155;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCCCTTGAA 21
Db 582 ATCGTTGCCCGCCCGCCCTTGAA 562

RESULT 25
US-09-491-577-106/c
; Sequence 106, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
; APPLICANT: Yale University
; APPLICANT: Carlson, John R.
; APPLICANT: Kim, Hunyong
; APPLICANT: Clyne, Peter J.
; APPLICANT: Warr, Coral G.
; TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
; FILE REFERENCE: 44574-5061-US
; CURRENT APPLICATION NUMBER: US/09/491,577
; CURRENT FILING DATE: 2000-01-25
; EARLIER APPLICATION NUMBER: US 60/117,132
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 1354
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: DOR 33B.3, region of BDGP Clone No. 6610511 AC006240
; NAME/KEY: intron
; LOCATION: (1)..(120)
; NAME/KEY: intron
; LOCATION: (852)..(930)
US-09-491-577-106

Query Match      73.6%; Score 16.2; DB 4; Length 1354;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCCCTTGAA 21
Db 702 ATCGTTGCCCGCCCGCCCTTGAA 682

RESULT 26
US-09-902-540-722
; Sequence 722, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B

; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 722
; LENGTH: 5787
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(5787)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-722

Query Match      73.6%; Score 16.2; DB 4; Length 5787;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCCCTTGAA 21
Db 3560 ATCCTTGCCCGCCCGCTTGAA 3580

RESULT 27
US-09-949-016-4205
; Sequence 4205, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4205
; LENGTH: 6882
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4205

Query Match      73.6%; Score 16.2; DB 4; Length 6882;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCGCCCTTGAA 21
Db 4560 ACCCATTCCCGCCCGCCCTTGAA 4580

RESULT 28
US-09-949-016-14299
; Sequence 14299, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```



```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14299
; LENGTH: 9391
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14299

```

Query Match 73.6%; Score 16.2; DB 4; Length 9391;
Best Local Similarity 85.7%; Pred. NO. 4.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0

Qy 1 ATCCTTGCCCCCGCCCTTGAA 21
423 ATCCTGGCCCGAGCCTCTTGAA 443
Db

RESULT 29
US-09-262-773-206
; Sequence 206. Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hesse, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1

```

; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 20137
; TYPE: DNA
; ORGANISM: human
US-09-262-773-206

```

Query Match 73.6%; Score 16.2; DB 3; Length 20137;
Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0;

QY 2 T C T T T G C C C C G C C C C T T G A A T 22
|||
Dp 1733 T C T T T G C C C C A G C A C C T T G A A T 1753

RESULT 30
US-09-262-773-9
US-09-262-773-9
; Sequence 9, Application US/09262773
; Patent NO. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hesse, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO: 3
; LENGTH: 20138
; TYPE: DNA
; ORGANISM: human
US-09-262-773-9

```

Query Match 73.6%; Score 16.2; DB 3; Length 20138;

Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 3

Qy 2 TCCTTGCCCCCGCCCTTGAAT 22
|||
Db 1734 TCTTTGCCACGACCTTGAAT 1754

```

RESULT 31
US-09-262-773-210
; Sequence 210, Application US/09262773
; Patent NO. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 210
; LENGTH: 23071
; TYPE: DNA
; ORGANISM: human
US-09-262-773-210

```

Query Match 73.6%; Score 16.2; DB 3; Length 23071;
Best Local Similarity 85.7%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0;

QY 2 TCCTTGCCCCGCCCCCTTGAAT 22

Db 4667 TCCTTGCCCAGCACCTTGAAT 4687

RESULT 32
US-09-949-016-14449/c
; Sequence 14449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14449
; LENGTH: 34088
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14449

```
Query Match      73.6%; Score 16.2; DB 4; Length 34088;
Best Local Similarity 85.7%; Pred. No. 5.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 2 TCCTTGCCCGCCCTTGAAT 22
| | | | | | | | | |
Dp 31328 TTCCTGCCCCTGCCCTTGAAT 31308

```
RESULT 33
US-09-949-016-13184
; Sequence 13184, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13184
; LENGTH: 51928
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(51928)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13184

Query Match      73.6%; Score 16.2; DB 4; Length 51928;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TCCTTGCCCGCCGCTTGAAT 22
DB      25762 TCCTTGCCCGCCGCTTGAAT 25782

RESULT 34
US-09-851-896-3
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; FILE REFERENCE: EXPRESSION
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match      73.6%; Score 16.2; DB 3; Length 70000;
Best Local Similarity 85.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TCCTTGCCCGCCGCTTGAAT 22
DB      17950 TCCTTGCCCGCCGCTTGAAT 17970

RESULT 35
US-09-949-016-16819
; Sequence 16819, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16819
; LENGTH: 76399
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(76399)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16819

Query Match      73.6%; Score 16.2; DB 4; Length 76399;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TCCTTGCCCGCCGCTTGAAT 22
DB      20993 TCCTTGCCCGCCGCTTGAAT 21013

RESULT 36
US-09-949-016-17375/c
; Sequence 17375, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17375
; LENGTH: 84296
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(84296)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17375

Query Match      73.6%; Score 16.2; DB 4; Length 84296;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATCCTTGCCCGCCGCTTGAAT 21
DB      3084 ATCCTTGCCCGCCGCTTCAA 3064

RESULT 37
US-09-949-016-15078/c
```

```
; Sequence 15078, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15078
; LENGTH: 114426
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15078

Query Match          73.6%; Score 16.2; DB 4; Length 114426;
Best Local Similarity 85.7%; Pred. No. 5.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
Db 28656 TCATTCCCGCCACCTTGAAT 28636

RESULT 38
US-09-949-016-12276/c
; Sequence 12276, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12276
; LENGTH: 171700
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12276

Query Match          73.6%; Score 16.2; DB 4; Length 171700;
Best Local Similarity 85.7%; Pred. No. 5.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
Db 112356 TCCTTGGCCCGCCCTTGAAT 112336

RESULT 39
US-09-949-016-15835/c
; Sequence 15835, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15835
; LENGTH: 171701
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15835

Query Match          73.6%; Score 16.2; DB 4; Length 171701;
Best Local Similarity 85.7%; Pred. No. 5.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
Db 112356 TCCTTGGCCCGCCCTTGAAT 112336

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15835
; LENGTH: 171701
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15835

Query Match          73.6%; Score 16.2; DB 4; Length 171701;
Best Local Similarity 85.7%; Pred. No. 5.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTTGGCCCGCCCTTGAAT 22
Db 112356 TCCTTGGCCCGCCCTTGAAT 112336

RESULT 40
US-09-949-016-17504
; Sequence 17504, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17504
; LENGTH: 276237
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(276237)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17504

Query Match          73.6%; Score 16.2; DB 4; Length 276237;
Best Local Similarity 85.7%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCCTTGGCCCGCCCTTGAAT 21
Db 95373 ATCCTTGGCCCGCCCTTGAAT 95393

RESULT 41
US-09-270-767-27636/c
; Sequence 27636, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
```

```
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27636
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27636

Query Match          71.8%; Score 15.8; DB 4; Length 740;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 CCTTGCCCGCCCTTGAA 21
      ||||| ||||| ||||| |||||
Db      247 CCTTGACCAGCCCTTGAA 229

RESULT 42
US-09-902-540-6737/c
; Sequence 6737, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6737
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6737

Query Match          71.8%; Score 15.8; DB 4; Length 858;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 CCTTGCCCGCCCTTGAA 21
      ||||| ||||| ||||| |||||
Db      91 CCATGCCCGCGCCCTTGAA 73

RESULT 43
US-09-902-540-9593/c
; Sequence 9593, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9593
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9593

Query Match          71.8%; Score 15.8; DB 4; Length 996;

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27636
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27636

Query Match          71.8%; Score 15.8; DB 4; Length 740;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 CCTTGCCCGCCCTTGAA 21
      ||||| ||||| ||||| |||||
Db      247 CCTTGACCAGCCCTTGAA 229

RESULT 44
US-09-270-767-11959/c
; Sequence 11959, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11959
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11959

Query Match          71.8%; Score 15.8; DB 4; Length 1540;
Best Local Similarity 89.5%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 CCTTGCCCGCCCTTGAA 21
      ||||| ||||| ||||| |||||
Db      1047 CCTTGACCAGCCCTTGAA 1029

RESULT 45
US-09-902-540-541/c
; Sequence 541, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 541
; LENGTH: 4009
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-541

Query Match          71.8%; Score 15.8; DB 4; Length 4009;
Best Local Similarity 89.5%; Pred. No. 6.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 CCTTGCCCGCCCTTGAA 21
      ||||| ||||| ||||| |||||
Db      2869 CCATGCCCGCCCTTGAA 2851

RESULT 46
US-09-902-540-1089
; Sequence 1089, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
```

```
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1089
; LENGTH: 13346
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1089

Query Match          71.8%; Score 15.8; DB 4; Length 13346;
Best Local Similarity 89.5%; Pred. No. 7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTTGCCTGGCCGCCCCCTTGAA 21
   ||| ||||| ||||| ||||| |||||
Db 6003 CCTGCGCCGCGCCCTTGAA 6021

RESULT 47
US-09-949-016-12533/c
; Sequence 12533, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12533
; LENGTH: 45275
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(45275)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12533

Query Match          71.8%; Score 15.8; DB 4; Length 45275;
Best Local Similarity 89.5%; Pred. No. 7.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTTGCCTGGCCGCCCCCTTGAA 21
   ||| ||||| ||||| ||||| |||||
Db 30576 CCTAGCCCGCCGCCCCCTTGAA 30558

RESULT 48
US-09-949-016-16198
; Sequence 16198, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1089
; LENGTH: 13346
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1089

Query Match          71.8%; Score 15.8; DB 4; Length 78810;
Best Local Similarity 89.5%; Pred. No. 8.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTTGCCTGGCCGCCCCCTTGAA 21
   ||| ||||| ||||| ||||| |||||
Db 29058 CCGTGCCTAGCCGCCCCCTTGAA 29076

RESULT 49
US-09-949-016-12298
; Sequence 12298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12298
; LENGTH: 115963
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(115963)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12298

Query Match          71.8%; Score 15.8; DB 4; Length 115963;
Best Local Similarity 89.5%; Pred. No. 8.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCTGGCCGCCCCCTTG 19
   ||||| ||||| ||||| |||||
Db 30330 ATCCTTGCCTGGCCGCCCCCTTG 30348

RESULT 50
US-09-949-016-15316
; Sequence 15316, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15316
; LENGTH: 144322
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(144322)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15316
```

```
Query Match          71.8%; Score 15.8; DB 4; Length 144322;
Best Local Similarity 89.5%; Pred.No. 8.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 ATCCTTGCCCGCCGCTTG 19
      ||||| ||||| |||||
Db      106689 ATCCTTGCCCGCCGCTTG 106707
```

```
Search completed: April 25, 2005, 15:10:40
Job time : 92 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:16:14 ; Search time 274.621 Seconds
(without alignments)
486.355 Million cell updates/sec

Title: US-10-010-476-14

Perfect score: 22

Sequence: 1 ATCTTGCCCGCCGCTTGAAAT 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	15	US-10-010-476-14
2	22	100.0	30	15	US-10-010-476-9
3	22	100.0	30	15	US-10-010-476-10
4	22	100.0	50	15	US-10-010-476-13
5	18.4	83.6	424	14	US-10-116-802-513
6	17.8	80.9	540	18	US-10-430-201-2349
7	17.8	80.9	540	17	US-10-430-201-2350
8	17.4	79.1	180	17	US-10-242-535A-6953
9	17.4	79.1	180	17	US-10-085-783A-6953
10	17.4	79.1	262	17	US-10-242-535A-42504
11	17.4	79.1	262	17	US-10-085-783A-42504

12	17.4	79.1	277	17	US-10-242-535A-9468	Sequence 9468, Ap
13	17.4	79.1	277	17	US-10-085-783A-9468	Sequence 9468, Ap
14	17.4	79.1	461	10	US-09-814-353-5737	Sequence 5737, Ap
15	17.4	79.1	461	10	US-09-814-353-12020	Sequence 12020, A
16	17.4	79.1	526	18	US-10-425-115-131927	Sequence 131927, A
17	17.4	79.1	1789	17	US-10-172-118-585	Sequence 585, App
18	17.4	79.1	1789	17	US-10-342-887-585	Sequence 585, App
19	17.2	78.2	574	18	US-10-425-115-170110	Sequence 170110, A
20	17.2	78.2	818	18	US-10-767-795-6234	Sequence 6234, Ap
21	17.2	78.2	191597	18	US-10-719-993-6802	Sequence 6802, Ap
22	17	77.3	1051	11	US-09-876-143-1396	Sequence 1396, Ap
23	17	77.3	2352	18	US-10-437-963-49030	Sequence 49030, A
24	16.8	76.4	577	13	US-10-027-632-16892	Sequence 16892, A
25	16.8	76.4	577	13	US-10-027-632-16893	Sequence 16893, A
26	16.8	76.4	577	13	US-10-027-632-16894	Sequence 16894, A
27	16.8	76.4	577	17	US-10-027-632-16892	Sequence 16892, A
28	16.8	76.4	577	17	US-10-027-632-16893	Sequence 16893, A
29	16.8	76.4	577	17	US-10-027-632-16894	Sequence 16894, A
30	16.8	76.4	596	10	US-09-764-891-9954	Sequence 9954, Ap
31	16.8	76.4	618	10	US-09-764-891-9955	Sequence 9955, Ap
32	16.8	76.4	621	10	US-09-764-891-9955	Sequence 9955, Ap
33	16.8	76.4	716	17	US-10-258-666-29	Sequence 29, Appl
34	16.8	76.4	787	17	US-10-258-666-30	Sequence 30, Appl
35	16.8	76.4	816	13	US-10-027-632-166755	Sequence 166755, A
36	16.8	76.4	816	13	US-10-027-632-166756	Sequence 166756, A
37	16.8	76.4	816	17	US-10-027-632-166755	Sequence 166755, A
38	16.8	76.4	816	17	US-10-027-632-166756	Sequence 166756, A
39	16.8	76.4	1202	13	US-10-027-632-101097	Sequence 101097, A
40	16.8	76.4	1202	17	US-10-027-632-101097	Sequence 101097, A
41	16.8	76.4	2818	18	US-10-437-963-41730	Sequence 41730, A
42	16.8	76.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
43	16.4	74.5	330	18	US-10-432-934-1	Sequence 1, Appl
44	16.4	74.5	505	18	US-10-425-115-86331	Sequence 86331, A
45	16.4	74.5	903	17	US-10-260-238-5715	Sequence 5715, Ap
46	16.4	74.5	1052	17	US-10-425-114-23849	Sequence 23849, A
47	16.4	74.5	1167	17	US-10-425-114-31322	Sequence 31322, A
48	16.4	74.5	1185	17	US-10-425-114-35735	Sequence 35735, A
49	16.4	74.5	1199	18	US-10-739-930-2611	Sequence 2611, Ap
50	16.4	74.5	1238	17	US-10-425-114-15826	Sequence 15826, A
51	16.4	74.5	1557	17	US-10-425-115-7566	Sequence 7566, Ap
52	16.4	74.5	1567	18	US-10-282-122A-23246	Sequence 23246, A
53	16.4	74.5	1933	18	US-10-425-115-38896	Sequence 38896, A
54	16.4	74.5	4739	14	US-10-176-667-1	Sequence 1, Appl
55	16.4	74.5	4739	14	US-10-176-640-1	Sequence 1, Appl
56	16.4	74.5	4739	14	US-10-176-495-1	Sequence 1, Appl
57	16.4	74.5	4739	14	US-10-178-495-1	Sequence 1, Appl
58	16.4	74.5	33454	13	US-10-087-192-1111	Sequence 1111, Ap
59	16.2	73.6	150	9	US-09-796-632-6134	Sequence 6134, Ap
60	16.2	73.6	150	14	US-10-040-862-6134	Sequence 6134, Ap
61	16.2	73.6	150	17	US-10-057-475B-6134	Sequence 6134, Ap
62	16.2	73.6	150	17	US-10-154-884B-6134	Sequence 6134, Ap
63	16.2	73.6	150	18	US-10-764-324-6134	Sequence 6134, Ap
64	16.2	73.6	155	16	US-10-029-386-24406	Sequence 24406, A
65	16.2	73.6	222	9	US-09-974-300-8149	Sequence 8149, Ap
66	16.2	73.6	257	16	US-10-029-386-22283	Sequence 22283, A
67	16.2	73.6	359	17	US-10-242-535A-2920	Sequence 2920, Ap
68	16.2	73.6	359	17	US-10-085-783A-2920	Sequence 2920, Ap
69	16.2	73.6	445	10	US-09-764-891-429	Sequence 429, App
70	16.2	73.6	445	15	US-10-205-428-61	Sequence 61, Appl
71	16.2	73.6	482	15	US-10-357-330-13565	Sequence 13565, A
72	16.2	73.6	508	16	US-10-029-386-10706	Sequence 10706, A
73	16.2	73.6	509	17	US-10-424-599-125632	Sequence 125632, A
74	16.2	73.6	545	16	US-10-029-386-8555	Sequence 8555, Ap
75	16.2	73.6	628	13	US-10-027-632-186030	Sequence 186030, A
76	16.2	73.6	628	17	US-10-027-632-186030	Sequence 186030, A
77	16.2	73.6	673	17	US-10-062-674-1732	Sequence 1732, Ap
78	16.2	73.6	689	13	US-10-027-632-215782	Sequence 215782, A
79	16.2	73.6	689	17	US-10-027-632-215782	Sequence 215782, A
80	16.2	73.6	752	13	US-10-027-632-28476	Sequence 28476, A
81	16.2	73.6	752	13	US-10-027-632-28477	Sequence 28477, A
82	16.2	73.6	752	13	US-10-027-632-28478	Sequence 28478, A
83	16.2	73.6	752	17	US-10-027-632-28476	Sequence 28476, A
84	16.2	73.6	752	17	US-10-027-632-28477	Sequence 28477, A

c 85	16.2	73.6	752	17	US-10-027-632-28478	Sequence 28478, A	15.8	71.8	1496	18	US-10-739-930-4379	Sequence 4379, Ap
c 86	16.2	73.6	762	17	US-10-369-493-24795	Sequence 24795, A	15.8	71.8	1708	18	US-10-425-115-154359	Sequence 154359, A
c 87	16.2	73.6	837	17	US-10-369-493-42868	Sequence 42868, A	15.8	71.8	1713	18	US-10-437-963-72391	Sequence 72391, A
c 88	16.2	73.6	1152	10	US-09-932-227-13	Sequence 13, Appl	15.8	71.8	1961	18	US-10-425-115-137382	Sequence 137382, A
c 89	16.2	73.6	1152	15	US-10-183-708-13	Sequence 13, Appl	15.8	71.8	2349	17	US-10-104-047-1148	Sequence 1148, Ap
c 90	16.2	73.6	1155	17	US-10-601-309-15	Sequence 15, Appl	15.8	71.8	2573	17	US-10-152-319A-2018	Sequence 2018, Ap
c 91	16.2	73.6	1354	17	US-10-601-309-106	Sequence 106, App	15.8	71.8	5070	18	US-10-437-963-60747	Sequence 60747, A
c 92	16.2	73.6	1419	17	US-10-369-493-32535	Sequence 32535, A	15.8	71.8	10024	15	US-10-017-161-1671	Sequence 1671, Ap
c 93	16.2	73.6	1787	17	US-10-104-047-1674	Sequence 1674, Ap	15.8	71.8	10102	17	US-10-232-798-1331	Sequence 1331, Ap
c 94	16.2	73.6	2073	18	US-10-425-115-172909	Sequence 172909, A	15.8	71.8	52637	17	US-10-367-094-186	Sequence 186, App
c 95	16.2	73.6	2149	17	US-10-108-260A-1865	Sequence 1865, Ap	15.8	71.8	94810	17	US-10-052-482-22	Sequence 22, Appl
c 96	16.2	73.6	2223	17	US-10-104-047-1664	Sequence 1664, Ap	15.8	71.8	96593	11	US-09-997-722-151	Sequence 151, App
c 97	16.2	73.6	2244	18	US-10-723-860-6634	Sequence 6634, Ap	15.8	71.8	344805	18	US-10-779-271-1	Sequence 1, Appl
c 98	16.2	73.6	2298	17	US-10-108-260A-1375	Sequence 1375, Ap	15.6	70.9	80	17	US-10-384-245-1010	Sequence 1010, Ap
c 99	16.2	73.6	2568	17	US-10-312-088-12	Sequence 12, Appl	15.6	70.9	93	14	US-10-060-036-3745	Sequence 3745, Ap
c 100	16.2	73.6	2622	15	US-10-176-306-81	Sequence 81, Appl	15.6	70.9	168	17	US-10-242-535A-30513	Sequence 30513, A
c 101	16.2	73.6	2634	15	US-10-176-306-79	Sequence 79, Appl	15.6	70.9	168	17	US-10-085-783A-30513	Sequence 30513, A
c 102	16.2	73.6	2667	17	US-10-312-088-11	Sequence 11, Appl	15.6	70.9	172	17	US-10-242-535A-31605	Sequence 31605, A
c 103	16.2	73.6	2913	18	US-10-437-963-77151	Sequence 77151, A	15.6	70.9	172	17	US-10-085-783A-31605	Sequence 31605, A
c 104	16.2	73.6	3657	14	US-10-198-846-14014	Sequence 14014, A	15.6	70.9	178	10	US-09-918-995-17933	Sequence 17933, A
c 105	16.2	73.6	5998	18	US-10-476-397-16	Sequence 16, Appl	15.6	70.9	201	19	US-10-741-600-33933	Sequence 33933, A
c 106	16.2	73.6	6591	18	US-10-723-860-7162	Sequence 7162, Ap	15.6	70.9	249	17	US-10-242-535A-40828	Sequence 40828, A
c 107	16.2	73.6	6952	17	US-10-221-613-209	Sequence 209, App	15.6	70.9	296	9	US-10-085-783A-40828	Sequence 40828, A
c 108	16.2	73.6	7008	19	US-10-482-834A-10	Sequence 10, Appl	15.6	70.9	296	17	US-09-964-824A-380	Sequence 380, App
c 109	16.2	73.6	7008	19	US-10-482-834A-16	Sequence 16, Appl	15.6	70.9	296	17	US-10-242-535A-27378	Sequence 27378, A
c 110	16.2	73.6	32816	9	US-09-729-094-3	Sequence 3, Appl	15.6	70.9	296	17	US-10-085-783A-27378	Sequence 27378, A
c 111	16.2	73.6	32816	16	US-10-435-631-3	Sequence 3, Appl	15.6	70.9	296	19	US-10-843-641A-5683	Sequence 5683, Ap
c 112	16.2	73.6	33916	18	US-10-322-281-861	Sequence 861, App	15.6	70.9	312	17	US-10-242-535A-19218	Sequence 19218, A
c 113	16.2	73.6	70000	17	US-10-476-991-3	Sequence 3, Appl	15.6	70.9	312	17	US-10-085-783A-19218	Sequence 19218, A
c 114	16.2	73.6	158001	18	US-10-211-179-11	GENERAL INFORMATI	15.6	70.9	322	14	US-10-060-036-2586	Sequence 2586, Ap
c 115	16.2	73.6	203132	18	US-10-322-281-459	Sequence 459, App	15.6	70.9	344	18	US-10-425-115-57237	Sequence 57237, A
c 116	16.2	73.6	430442	18	US-10-417-375-128	Sequence 128, App	15.6	70.9	345	17	US-10-242-535A-51379	Sequence 51379, A
c 117	16.2	73.6	3673778	16	US-10-312-841-1	Sequence 1, Appl	15.6	70.9	345	17	US-10-085-783A-51379	Sequence 51379, A
c 118	16	72.7	326	10	US-09-803-719-2383	Sequence 2383, Ap	15.6	70.9	368	14	US-10-060-036-3437	Sequence 3437, Ap
c 119	16	72.7	340	10	US-09-803-719-2387	Sequence 2387, Ap	15.6	70.9	371	18	US-10-425-115-97766	Sequence 97766, A
c 120	16	72.7	13273	9	US-09-764-869-2349	Sequence 2349, Ap	15.6	70.9	383	18	US-10-357-930-17232	Sequence 17232, A
c 121	16	72.7	13273	14	US-10-091-504-2349	Sequence 2349, Ap	15.6	70.9	403	18	US-10-425-115-78839	Sequence 78839, A
c 122	16	72.7	13273	17	US-10-227-577-2349	Sequence 2349, Ap	15.6	70.9	433	17	US-10-242-535A-26249	Sequence 26249, A
c 123	15.8	71.8	365	18	US-10-425-115-114370	Sequence 114370, A	15.6	70.9	433	17	US-10-085-783A-26249	Sequence 26249, A
c 124	15.8	71.8	384	18	US-10-425-115-29431	Sequence 29431, A	15.6	70.9	472	17	US-10-424-599-104460	Sequence 104460, A
c 125	15.8	71.8	396	17	US-10-242-535A-9226	Sequence 9226, Ap	15.6	70.9	473	18	US-10-425-115-23260	Sequence 23260, A
c 126	15.8	71.8	396	17	US-10-085-783A-9226	Sequence 9226, Ap	15.6	70.9	474	18	US-10-425-115-23260	Sequence 23260, A
c 127	15.8	71.8	453	9	US-09-864-761-5078	Sequence 5078, Ap	15.6	70.9	482	17	US-10-242-535A-41870	Sequence 41870, A
c 128	15.8	71.8	456	9	US-09-783-590-3630	Sequence 3630, Ap	15.6	70.9	482	17	US-10-085-783A-41870	Sequence 41870, A
c 129	15.8	71.8	472	9	US-09-864-761-4428	Sequence 4428, Ap	15.6	70.9	490	13	US-10-027-632-85090	Sequence 85090, A
c 130	15.8	71.8	478	18	US-10-437-963-63902	Sequence 63902, A	15.6	70.9	490	13	US-10-027-632-85091	Sequence 85091, A
c 131	15.8	71.8	525	18	US-10-425-115-164024	Sequence 164024, A	15.6	70.9	490	17	US-10-027-632-85090	Sequence 85090, A
c 132	15.8	71.8	579	17	US-10-425-114-3052	Sequence 3052, Ap	15.6	70.9	490	17	US-10-027-632-85091	Sequence 85091, A
c 133	15.8	71.8	598	13	US-10-027-632-292848	Sequence 292848, A	15.6	70.9	503	18	US-10-437-963-38834	Sequence 38834, A
c 134	15.8	71.8	598	17	US-10-027-632-292848	Sequence 292848, A	15.6	70.9	510	13	US-10-027-632-39157	Sequence 39157, A
c 135	15.8	71.8	609	13	US-10-027-632-41010	Sequence 41010, A	15.6	70.9	510	13	US-10-027-632-72152	Sequence 72152, A
c 136	15.8	71.8	609	17	US-10-027-632-41010	Sequence 41010, A	15.6	70.9	510	13	US-10-027-632-303079	Sequence 303079, A
c 137	15.8	71.8	625	13	US-10-027-632-65158	Sequence 65158, A	15.6	70.9	510	17	US-10-027-632-39157	Sequence 39157, A
c 138	15.8	71.8	625	13	US-10-027-632-65158	Sequence 65158, A	15.6	70.9	510	17	US-10-027-632-72152	Sequence 72152, A
c 139	15.8	71.8	625	13	US-10-027-632-294409	Sequence 294409, A	15.6	70.9	510	17	US-10-027-632-303079	Sequence 303079, A
c 140	15.8	71.8	625	13	US-10-027-632-297232	Sequence 297232, A	15.6	70.9	524	18	US-10-357-930-43103	Sequence 43103, A
c 141	15.8	71.8	625	17	US-10-027-632-65158	Sequence 65158, A	15.6	70.9	534	18	US-10-357-930-47055	Sequence 47055, A
c 142	15.8	71.8	625	17	US-10-027-632-65159	Sequence 65159, A	15.6	70.9	542	13	US-10-027-632-91111	Sequence 91111, A
c 143	15.8	71.8	625	17	US-10-027-632-294409	Sequence 294409, A	15.6	70.9	542	13	US-10-027-632-91111	Sequence 91111, A
c 144	15.8	71.8	734	17	US-10-027-632-297232	Sequence 297232, A	15.6	70.9	542	13	US-10-027-632-304379	Sequence 304379, A
c 145	15.8	71.8	734	17	US-10-260-238-4268	Sequence 4268, Ap	15.6	70.9	542	13	US-10-027-632-304379	Sequence 304379, A
c 146	15.8	71.8	799	18	US-10-437-963-30515	Sequence 30515, A	15.6	70.9	542	17	US-10-027-632-91111	Sequence 91111, A
c 147	15.8	71.8	813	13	US-10-027-632-38465	Sequence 38465, A	15.6	70.9	542	17	US-10-027-632-91111	Sequence 91111, A
c 148	15.8	71.8	813	13	US-10-027-632-38466	Sequence 38466, A	15.6	70.9	542	17	US-10-027-632-304378	Sequence 304378, A
c 149	15.8	71.8	813	17	US-10-027-632-38465	Sequence 38465, A	15.6	70.9	542	17	US-10-027-632-304379	Sequence 304379, A
c 150	15.8	71.8	813	17	US-10-027-632-38465	Sequence 38466, A	15.6	70.9	553	13	US-10-027-632-288262	Sequence 288262, A
c 151	15.8	71.8	818	13	US-10-027-632-153893	Sequence 153893, A	15.6	70.9	553	13	US-10-027-632-288262	Sequence 288262, A
c 152	15.8	71.8	818	17	US-10-027-632-153893	Sequence 153893, A	15.6	70.9	553	13	US-10-027-632-288262	Sequence 288262, A
c 153	15.8	71.8	861	18	US-10-739-930-4881	Sequence 4881, Ap	15.6	70.9	553	17	US-10-027-632-288263	Sequence 288263, A
c 154	15.8	71.8	1008	11	US-09-876-143-1543	Sequence 1543, Ap	15.6	70.9	571	17	US-10-027-632-135325	Sequence 135325, A
c 155	15.8	71.8	1155	18	US-10-437-963-50210	Sequence 50210, Ap	15.6	70.9	571	17	US-10-027-632-135325	Sequence 135325, A
c 156	15.8	71.8	1299	17	US-10-282-122A-13843	Sequence 13843, A	15.6	70.9	628	13	US-10-027-632-246670	Sequence 246670, A
c 157	15.8	71.8	1344	17	US-10-425-114-12309	Sequence 12309, A	15.6	70.9	628	17	US-10-027-632-246670	Sequence 246670, A

REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-010-476-9

Query Match 100.0%; Score 22; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGAT 22
DB 26 ATCCTTGCCCGCCCTTGAT 5
|||||

RESULT 3
US-10-010-476-10
; Sequence 10, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402Albert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; METHYLTRANSFERASE AND METHODS FOR USE THEREOF
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-010-476-10

Query Match 100.0%; Score 22; DB 15; Length 30;

```
; Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGAAT 22
Db |||||||
5 ATCCTTGCCCGCCCTTGAAT 26

RESULT 4
US-10-010-476-13
; Sequence 13, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-010-476-13

Query Match 100.0%; Score 22; DB 15; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGAAT 22
Db |||||||
15 ATCCTTGCCCGCCCTTGAAT 36

RESULT 5
US-10-116-802-513/c
; Sequence 513, Application US/10116802
; Publication No. US20030065157A1
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US

; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 513
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID NO: 902956.2
US-10-116-802-513

Query Match 83.6%; Score 18.4; DB 14; Length 424;
Best Local Similarity 95.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGCCCGCCCTTGAAT 22
Db |||||||
343 CGGTGCCCGCCCTTGAAT 324

RESULT 6
US-10-430-201-2349
; Sequence 2349, Application US/10430201
; Publication No. US20040162679A1
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment
; FILE REFERENCE: 40716 (IP-010)
; CURRENT APPLICATION NUMBER: US/10/430,201
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/370,114
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 4879
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2349
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: {384}..(399)
; OTHER INFORMATION: n is a, c, g, or t
US-10-430-201-2349

Query Match 80.9%; Score 17.8; DB 18; Length 540;
Best Local Similarity 90.5%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db |||||||
280 TCCTTGCCCGCCCTTGAAT 300

RESULT 7
US-10-430-201-2350
; Sequence 2350, Application US/10430201
; Publication No. US20040162679A1
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment
; FILE REFERENCE: 40716 (IP-010)
; CURRENT APPLICATION NUMBER: US/10/430,201
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/370,114
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 4879
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2350
; LENGTH: 540
```

```
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (384)..(399)
; OTHER INFORMATION: n is a, c, g, or t
US-10-430-201-2350

Query Match      80.9%; Score 17.8; DB 18; Length 540;
Best Local Similarity 90.5%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAAT 22
Db 280 TCCTTCCCTGCCCTTGAAT 300

RESULT 8
US-10-242-535A-6953
; Sequence 6953, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6953
; TYPE: DNA
; LENGTH: 180
; ORGANISM: Human
US-10-242-535A-6953

Query Match      79.1%; Score 17.4; DB 17; Length 180;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTG 19
Db 6 ATCCTTGCCCGCCCTTG 24

RESULT 9
US-10-085-783A-6953
; Sequence 6953, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6953
```

```
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-6953

Query Match      79.1%; Score 17.4; DB 17; Length 180;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTG 19
Db 6 ATCCTTGCCCGCCCTTG 24

RESULT 10
US-10-242-535A-42504
; Sequence 42504, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42504
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11)..(11)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-42504

Query Match      79.1%; Score 17.4; DB 17; Length 262;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTG 19
Db 98 ATCCTTGCCCGCCCTTG 116

RESULT 11
US-10-085-783A-42504
; Sequence 42504, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42504
```

```
; SEQ ID NO 42504
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (11)..(11)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-42504

Query Match          79.1%; Score 17.4; DB 17; Length 262;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTG 19
|||||
Db 98 ATCCTTGCCCGCCCTTG 116
|||||

RESULT 12
US-10-242-535A-9468
; Sequence 9468, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US 10/085,783
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9468
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (37)..(37)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (62)..(62)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (68)..(68)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-9468

Query Match          79.1%; Score 17.4; DB 17; Length 277;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTTGCCCGCCCTTG 19
|||||
Db 116 ATCCTTGCCCGCCCTTG 134
|||||

RESULT 14
US-09-814-353-5737/c
; Sequence 5737, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
```

```

; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5737
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 90, 138, 152, 154, 166, 167, 186, 193, 208, 212, 214, 226,
; LOCATION: 243, 246, 266, 272, 285, 299, 306, 312, 315, 319, 327, 345,
; LOCATION: 351, 353, 355, 367, 369, 371, 393, 401, 416, 417, 422, 423,
; LOCATION: 434, 446, 452
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-5737

Query Match          79.1%; Score 17.4; DB 10; Length 461;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAA 21
   ||| ||||| ||||| |||||
DB 133 CCTTCCCGCGCCCTTGAA 115

RESULT 15
US-09-814-353-12020/c
; Sequence 12020, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-0068
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12020
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 90, 138, 152, 154, 166, 167, 186, 193, 208, 212, 214, 226,
; LOCATION: 243, 246, 266, 272, 285, 299, 306, 312, 315, 319, 327, 345,
; LOCATION: 351, 353, 355, 367, 369, 371, 393, 401, 416, 417, 422, 423,
; LOCATION: 434, 446, 452
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-12020

Query Match          79.1%; Score 17.4; DB 10; Length 461;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAA 21
   ||| ||||| ||||| |||||
DB 133 CCTTCCCGCGCCCTTGAA 115

RESULT 16
US-10-425-115-131927
; Sequence 131927, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 131927
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_517C.1
US-10-425-115-131927

Query Match          79.1%; Score 17.4; DB 18; Length 526;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTGGCCCGCCCTTGAA 21
   ||| ||||| ||||| |||||
DB 360 CCTTGGCCCGCCCTTTAA 378

RESULT 17
US-10-172-118-585
; Sequence 585, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 585
; LENGTH: 1789
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_001267
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-585

Query Match          79.1%; Score 17.4; DB 17; Length 1789;
Best Local Similarity 94.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGGCCCGCCCTTG 19
   ||| ||||| ||||| |||||
DB 1622 ATCCTTGGCCCGCCCTTG 1640

RESULT 18
```



```
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 1744
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1396
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1051)
; OTHER INFORMATION: n is a, c, g, or t
US-09-876-143-1396

Query Match          77.3%; Score 17; DB 11; Length 1051;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCCTGCCCCGCCCTTGAA 21
Db 752 TCCTGCCCCCNCCCTTAA 733

RESULT 23
US-10-437-963-49030/c
; Sequence 49030, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 49030
; LENGTH: 2352
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_5164C.1
US-10-437-963-49030

Query Match          77.3%; Score 17; DB 18; Length 2352;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTTGCCCCGCCCTTG 19
Db 787 CCTTGCCCCGCCCTTG 771

RESULT 24
US-10-027-632-16892/c
; Sequence 16892, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; OTHER INFORMATION: Cloning of Single Nucleotide
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16892
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16892

Query Match          76.4%; Score 16.8; DB 13; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCTTGAA 21
Db 265 TCCTTGCCCCGCCCTTGGA 246

RESULT 25
US-10-027-632-16893/c
; Sequence 16893, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16893
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16893

Query Match          76.4%; Score 16.8; DB 13; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCCGCCCTTGAA 21
Db 265 TCCTTGCCCCGCCCTTGGA 246

RESULT 26
US-10-027-632-16894/c
; Sequence 16894, Application US/10027632
; Publication No. US20020198371A1
```

GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16894
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16894

Query Match 76.4%; Score 16.8; DB 13; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 265 TCCTTGCCAGCCCTTGGA 246

RESULT 27
US-10-027-632-16892/c
; Sequence 16892, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16892
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16892

Query Match 76.4%; Score 16.8; DB 17; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TCCTTGCCCGCCCTTGAA 21
DB 265 TCCTTGCCAGCCCTTGGA 246

RESULT 28
US-10-027-632-16893/c
; Sequence 16893, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16893
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16893

Query Match 76.4%; Score 16.8; DB 17; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 265 TCCTTGCCAGCCCTTGGA 246

RESULT 29
US-10-027-632-16894/c
; Sequence 16894, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

```
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16894
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-16894

Query Match          76.4%; Score 16.8; DB 17; Length 577;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 265 TCCTTGCCCGCCCTTGGA 246

RESULT 30
US-09-764-891-9954
; Sequence 9954, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9954
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9954

Query Match          76.4%; Score 16.8; DB 10; Length 596;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 111 TCCTTGCCCTGGCCCTTGAA 130

RESULT 31
US-09-764-891-9956
; Sequence 9956, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9956
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9956

Query Match          76.4%; Score 16.8; DB 10; Length 618;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 133 TCCTTGCCCTGGCCCTTGAA 152

RESULT 32
US-09-764-891-9955
; Sequence 9955, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9955
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9955

Query Match          76.4%; Score 16.8; DB 10; Length 621;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
DB 136 TCCTTGCCCTGGCCCTTGAA 155

RESULT 33
US-10-258-666-29/c
; Sequence 29, Application US/10258666
; Publication No. US20040005578A1
; GENERAL INFORMATION:
; APPLICANT: Yamada, Yoji
; APPLICANT: Sekine, Susumu
; APPLICANT: Kikuchi, Yasuhiro
; APPLICANT: Kyowa Hakko Kogyo Co., Ltd.
; TITLE OF INVENTION: Myocardial Cell Proliferation-Associated Genes
; FILE REFERENCE: 082382-000000US
; CURRENT APPLICATION NUMBER: US/10/258,666
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: JP 2000-126741
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: WO PCT/JP01/03700
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: clone RHDH-279, homologue of mouse interferon
; OTHER INFORMATION: regulatory factor 3 (mirf3)
US-10-258-666-29

Query Match          76.4%; Score 16.8; DB 17; Length 716;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCTTGCCCGCCCTTGAA 22
DB 166 CCTTGCCCGCCCTTGAGT 147

RESULT 34
US-10-258-666-30/c
; Sequence 30, Application US/10258666
; Publication No. US20040005578A1
; GENERAL INFORMATION:
; APPLICANT: Yamada, Yoji
```

; APPLICANT: Sekine, Susumu
; APPLICANT: Kikuchi, Yasuhiro
; APPLICANT: Sakurada, Kazuhiro
; APPLICANT: Kyowa Hakko Kogyo Co., Ltd.
; TITLE OF INVENTION: Myocardial Cell Proliferation-Associated Genes
; FILE REFERENCE: 082382-000000US
; CURRENT APPLICATION NUMBER: US/10/258,666
; PRIOR FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: JP 2000-126741
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: WO PCT/JP01/03700
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 787
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: clone RHDH-279-1, homologue of mouse interferon
; OTHER INFORMATION: regulatory factor 3 (mirf3)
; NAME/KEY: CDS
; LOCATION: (423)..(707)
; OTHER INFORMATION: RHDH-279-1 ORF
US-10-258-666-30

Query Match 76.4%; Score 16.8; DB 17; Length 787;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CCTTGGCCCGCCCTTGAA 22
|||||
Db 237 CCTTGGCCCGCCCTTGAG 218

RESULT 35
US-10-027-632-166755
; Sequence 166755, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166755
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166755

Query Match 76.4%; Score 16.8; DB 13; Length 816;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAA 21
|||||
Db 532 TCCTTGCCCGCCCTTGAA 551

RESULT 36
US-10-027-632-166756
; Sequence 166756, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166756
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166756

Query Match 76.4%; Score 16.8; DB 13; Length 816;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCCTTGCCCGCCCTTGAA 21
|||||
Db 532 TCCTTGCCCGCCCTTGAA 551

RESULT 37
US-10-027-632-166755
; Sequence 166755, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166755
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166755

Query Match
Best Local Similarity 76.4%; Score 16.8; DB 17; Length 816;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
| | | | | | | | | | | | | | | | | | | |
Db 532 TCCTTGCCCGCCCTTGAA 551

RESULT 38
US-10-027-632-166756
; Sequence 166756, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166756
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166756

Query Match
Best Local Similarity 76.4%; Score 16.8; DB 17; Length 816;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
| | | | | | | | | | | | | | | | | | | |
Db 532 TCCTTGCCCGCCCTTGAA 551

RESULT 39
US-10-027-632-101097/c
; Sequence 101097, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166756
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166756

Query Match
Best Local Similarity 76.4%; Score 16.8; DB 17; Length 816;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
| | | | | | | | | | | | | | | | | | | |
Db 532 TCCTTGCCCGCCCTTGAA 551

RESULT 39
US-10-027-632-101097/c
; Sequence 101097, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101097
; LENGTH: 1202
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101097

Query Match
Best Local Similarity 76.4%; Score 16.8; DB 17; Length 1202;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
| | | | | | | | | | | | | | | | | | | |
Db 412 TCCTTGCCCGCCCTTGTA 393

RESULT 40
US-10-027-632-101097/c
; Sequence 101097, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101097
; LENGTH: 1202
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101097

Query Match
Best Local Similarity 76.4%; Score 16.8; DB 17; Length 1202;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGAA 21
| | | | | | | | | | | | | | | | | | | |
Db 412 TCCTTGCCCGCCCTTGTA 393

RESULT 41
US-10-437-963-41730/c
; Sequence 41730, Application US/10437963
; Publication No. US20040123343A1
```

GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 41730
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_45049C.1
US-10-437-963-41730

Query Match 76.4%; Score 16.8; DB 18; Length 2818;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTTGA 20
|||||
DB 1843 ATCCTTGCCCGCCGCTGA 1824

RESULT 42

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 76.4%; Score 16.8; DB 15; Length 9025608;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTGA 21
|||||
DB 6122668 TCCTTGCCCGCCCGTGA 6122687

RESULT 43

US-10-432-934-1/c
; Sequence 1, Application US/10432934
; Publication No. US20040110269A1
; GENERAL INFORMATION:
; APPLICANT: Vipond, Richard
; APPLICANT: Shuttleworth, Helen
; APPLICANT: Ambrose, Emma
; APPLICANT: Minton, Nigel Peter
; TITLE OF INVENTION: Protection against mycobacterial infections
; FILE REFERENCE: 1581.0970001
; CURRENT APPLICATION NUMBER: US/10/432,934
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: PCT/GB01/05250
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: GB0028966.0
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-432-934-1

Query Match 74.5%; Score 16.4; DB 18; Length 330;
Best Local Similarity 94.4%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
|||||
DB 247 TCCTTGCCCGACCTTG 230

RESULT 44

US-10-425-115-86331/c
; Sequence 86331, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 86331
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_178743C.1
US-10-425-115-86331

Query Match 74.5%; Score 16.4; DB 18; Length 505;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCTTGCCCGCCCTT 18
|||||
DB 479 ATCCTTGCCCGCCCTT 462

RESULT 45

US-10-260-238-5715
; Sequence 5715, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.

```

; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, FumiYaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricks, Barrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 5715
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Zea mays
US-10-260-238-5715

Query Match          74.5%; Score 16.4; DB 17; Length 903;
Best Local Similarity 94.4%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
DB 62 TCCTTGCCCGCCCTCG 79

RESULT 46
US-10-425-114-23849
; Sequence 23849, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23849
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB36-020-B8_FLI
US-10-425-114-23849

Query Match          74.5%; Score 16.4; DB 17; Length 1052;
Best Local Similarity 94.4%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
DB 82 TCCTTGCCCGCCCTCG 99

RESULT 47
US-10-425-114-31322
; Sequence 31322, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15735
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-063-D6_FLI
US-10-425-114-15735

Query Match          74.5%; Score 16.4; DB 17; Length 1185;
Best Local Similarity 94.4%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
DB 194 TCCTTGCCCGCCCTCG 211

RESULT 48
US-10-425-114-15735
; Sequence 15735, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15735
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-063-D6_FLI
US-10-425-114-15735

Query Match          74.5%; Score 16.4; DB 17; Length 1185;
Best Local Similarity 94.4%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTTGCCCGCCCTTG 19
DB 194 TCCTTGCCCGCCCTCG 211

RESULT 49
US-10-739-930-2611
; Sequence 2611, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:14:29 ; Search time 42 Seconds
(without alignments)
545.426 Million cell updates/sec

Title: US-10-010-476-15
Perfect score: 14
Sequence: 1 TTGCCCCCCCCCTT 14

Scoring table: IDENTITY_NUC
Gapop 10_0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS-COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	14	100.0	26700	1	US-08-472-217-1
C 2	14	100.0	26700	2	US-08-488-199-5
C 3	14	100.0	26700	3	US-08-760-534A-1
C 4	14	100.0	26700	4	US-09-336-757-1
C 5	13	92.9	509	4	US-09-270-767-6481
C 6	13	92.9	509	4	US-09-270-767-21763
C 7	13	92.9	601	4	US-09-949-016-21120
C 8	13	92.9	601	4	US-09-949-016-47488
C 9	13	92.9	601	4	US-09-949-016-50128
C 10	13	92.9	601	4	US-09-949-016-50129
C 11	13	92.9	601	4	US-09-949-016-94941
C 12	13	92.9	609	4	US-09-621-976-1968
C 13	13	92.9	794	3	US-08-998-416-459
C 14	13	92.9	831	4	US-09-522-991A-3585
C 15	13	92.9	1195	4	US-09-522-714-7
C 16	13	92.9	1326	4	US-09-620-312D-867
C 17	13	92.9	1409	6	5242798-6
C 18	13	92.9	1409	6	5242798-6
C 19	13	92.9	1910	3	US-09-149-476-297
C 20	13	92.9	1945	3	US-09-149-476-164
C 21	13	92.9	2058	4	US-09-489-039A-5810
C 22	13	92.9	4812	4	US-09-949-016-17058
C 23	13	92.9	6821	4	US-09-902-540-907
C 24	13	92.9	7846	4	US-09-949-016-14510
C 25	13	92.9	8641	4	US-09-949-016-14077
C 26	13	92.9	9620	4	US-09-949-016-13551
C 27	13	92.9	11231	4	US-09-949-016-14030
C 28	13	92.9	12163	4	US-09-949-016-15030
C 29	13	92.9	12164	4	US-09-949-016-12202
C 30	13	92.9	15585	4	US-09-949-016-11927
C 31	13	92.9	15585	4	US-09-949-016-15627
C 32	13	92.9	31467	4	US-09-949-016-13134
C 33	13	92.9	31868	4	US-09-949-016-11907
C 34	13	92.9	33748	4	US-09-949-016-14283
C 35	13	92.9	41768	4	US-09-902-540-1266
C 36	13	92.9	53806	4	US-09-949-016-12572
C 37	13	92.9	83210	4	US-09-949-016-14209
C 38	13	92.9	97989	4	US-09-949-016-13208
C 39	13	92.9	101011	4	US-09-949-016-16933
C 40	13	92.9	4403765	3	US-09-103-840A-2
C 41	13	92.9	4411529	3	US-09-103-840A-1
C 42	12.4	88.6	39	3	US-08-556-978B-34
C 43	12.4	88.6	69	2	US-08-343-443B-46
C 44	12.4	88.6	94	4	US-09-626-002-28
C 45	12.4	88.6	95	4	US-09-626-002-27
C 46	12.4	88.6	113	4	US-09-626-002-25
C 47	12.4	88.6	137	4	US-09-626-002-24
C 48	12.4	88.6	153	4	US-09-626-002-20
C 49	12.4	88.6	202	4	US-09-513-999C-14750
C 50	12.4	88.6	209	4	US-09-626-002-31
C 51	12.4	88.6	210	4	US-09-248-796A-7453
C 52	12.4	88.6	238	4	US-09-016-434-518
C 53	12.4	88.6	251	4	US-09-513-999C-11704
C 54	12.4	88.6	273	4	US-09-328-352-15
C 55	12.4	88.6	294	4	US-09-489-039A-5390
C 56	12.4	88.6	296	4	US-09-626-002-23
C 57	12.4	88.6	296	4	US-09-626-002-29
C 58	12.4	88.6	303	3	US-08-556-978B-80
C 59	12.4	88.6	318	4	US-09-513-999C-11831
C 60	12.4	88.6	354	4	US-09-902-540-2404
C 61	12.4	88.6	375	2	US-08-793-107-7
C 62	12.4	88.6	375	3	US-08-793-110-7
C 63	12.4	88.6	375	3	US-08-793-109-1
C 64	12.4	88.6	408	4	US-09-513-999C-473
C 65	12.4	88.6	408	4	US-09-471-276-150
C 66	12.4	88.6	418	4	US-09-513-999C-3338
C 67	12.4	88.6	445	4	US-09-626-002-22
C 68	12.4	88.6	467	4	US-09-621-976-18116
C 69	12.4	88.6	472	4	US-09-596-248D-3
C 70	12.4	88.6	493	4	US-09-621-976-18630
C 71	12.4	88.6	502	4	US-09-626-002-20
C 72	12.4	88.6	506	4	US-09-513-999C-3966
C 73	12.4	88.6	510	4	US-09-626-002-3
C 74	12.4	88.6	536	4	US-09-621-976-1160
C 75	12.4	88.6	595	3	US-09-328-111-216
C 76	12.4	88.6	601	4	US-09-949-016-23046
C 77	12.4	88.6	601	4	US-09-949-016-23047
C 78	12.4	88.6	601	4	US-09-949-016-23048
C 79	12.4	88.6	601	4	US-09-949-016-23048
C 80	12.4	88.6	601	4	US-09-949-016-46319
C 81	12.4	88.6	601	4	US-09-949-016-46320
C 82	12.4	88.6	601	4	US-09-949-016-52316
C 83	12.4	88.6	601	4	US-09-949-016-52316
C 84	12.4	88.6	601	4	US-09-949-016-57331
C 85	12.4	88.6	601	4	US-09-949-016-88482
C 86	12.4	88.6	601	4	US-09-949-016-112415
C 87	12.4	88.6	601	4	US-09-949-016-125370
C 88	12.4	88.6	601	4	US-09-949-016-125371
C 89	12.4	88.6	601	4	US-09-949-016-171938
C 90	12.4	88.6	601	4	US-09-949-016-180659
C 91	12.4	88.6	601	4	US-09-949-016-200518
C 92	12.4	88.6	601	4	US-09-949-016-200519
C 93	12.4	88.6	601	4	US-09-949-016-200520
C 94	12.4	88.6	621	3	US-08-935-433-3
C 95	12.4	88.6	621	3	US-09-553-133-3
C 96	12.4	88.6	658	4	US-09-626-002-19
C 97	12.4	88.6	687	4	US-09-596-248D-66
C 98	12.4	88.6	738	4	US-09-596-248D-64
C 99	12.4	88.6	741	4	US-09-621-976-3446
C 100	12.4	88.6	762	4	US-09-270-767-1787

Sequence 16030, A
Sequence 12202, A
Sequence 11927, A
Sequence 15637, A
Sequence 13134, A
Sequence 11907, A
Sequence 14283, A
Sequence 1265, Ap
Sequence 12572, A
Sequence 14209, A
Sequence 13208, A
Sequence 16933, A
Sequence 2, Appli
Sequence 1, Appli
Sequence 34, Appli
Sequence 46, Appli
Sequence 28, Appli
Sequence 27, Appli
Sequence 26, Appli
Sequence 25, Appli
Sequence 24, Appli
Sequence 30, Appli
Sequence 14750, A
Sequence 31, Appli
Sequence 7453, Ap
Sequence 518, App
Sequence 11704, A
Sequence 15, Appli
Sequence 5390, Ap
Sequence 23, Appli
Sequence 29, Appli
Sequence 80, Appli
Sequence 11831, A
Sequence 2404, Ap
Sequence 7, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 473, App
Sequence 150, App
Sequence 3338, Ap
Sequence 22, Appli
Sequence 18116, A
Sequence 3, Appli
Sequence 18630, A
Sequence 20, Appli
Sequence 3966, Ap
Sequence 3, Appli
Sequence 1160, Ap
Sequence 216, App
Sequence 23046, A
Sequence 23047, A
Sequence 23048, A
Sequence 32393, A
Sequence 46319, A
Sequence 46320, A
Sequence 52216, A
Sequence 57331, A
Sequence 88482, A
Sequence 112415, A
Sequence 125370, A
Sequence 125371, A
Sequence 171938, A
Sequence 180659, A
Sequence 200518, A
Sequence 200519, A
Sequence 200520, A
Sequence 3, Appli
Sequence 3, Appli
Sequence 19, Appli
Sequence 66, Appli
Sequence 64, Appli
Sequence 3446, Ap
Sequence 1787, Ap

c 247	12	85.7	22	2	US-08-753-979A-10	Sequence 10, Appl	320	12	85.7	601	4	US-09-949-016-104305	Sequence 104305,
c 248	12	85.7	22	2	US-08-753-979A-24	Sequence 24, Appl	321	12	85.7	601	4	US-09-949-016-104375	Sequence 104375,
c 249	12	85.7	22	2	US-08-753-979A-40	Sequence 40, Appl	322	12	85.7	601	4	US-09-949-016-104376	Sequence 104376,
c 250	12	85.7	29	3	US-08-726-807B-36	Sequence 36, Appl	323	12	85.7	601	4	US-09-949-016-104446	Sequence 104446,
c 251	12	85.7	29	3	US-09-258-367-36	Sequence 36, Appl	324	12	85.7	601	4	US-09-949-016-104447	Sequence 104447,
c 252	12	85.7	29	3	US-09-546-550-36	Sequence 36, Appl	325	12	85.7	601	4	US-09-949-016-104517	Sequence 104517,
c 253	12	85.7	29	3	US-09-431-414-36	Sequence 36, Appl	326	12	85.7	601	4	US-09-949-016-104518	Sequence 104518,
c 254	12	85.7	29	3	US-09-225-670-36	Sequence 36, Appl	327	12	85.7	601	4	US-09-949-016-104588	Sequence 104588,
c 255	12	85.7	29	3	US-09-431-349C-36	Sequence 36, Appl	328	12	85.7	601	4	US-09-949-016-104589	Sequence 104589,
c 256	12	85.7	47	1	US-08-171-389-59	Sequence 59, Appl	329	12	85.7	601	4	US-09-949-016-104659	Sequence 104659,
c 257	12	85.7	47	1	US-08-123-936-59	Sequence 59, Appl	330	12	85.7	601	4	US-09-949-016-104660	Sequence 104660,
c 258	12	85.7	47	2	US-08-475-228A-59	Sequence 59, Appl	331	12	85.7	601	4	US-09-949-016-104730	Sequence 104730,
c 259	12	85.7	47	3	US-08-482-080A-59	Sequence 59, Appl	332	12	85.7	601	4	US-09-949-016-104731	Sequence 104731,
c 260	12	85.7	47	3	US-09-354-947-59	Sequence 59, Appl	c 333	12	85.7	601	4	US-09-949-016-105200	Sequence 105200,
c 261	12	85.7	47	5	US-09-131-12388-59	Sequence 59, Appl	c 334	12	85.7	601	4	US-09-949-016-110084	Sequence 110084,
c 262	12	85.7	51	3	US-08-863-813A-60	Sequence 60, Appl	c 335	12	85.7	601	4	US-09-949-016-110085	Sequence 110085,
c 263	12	85.7	51	4	US-09-500-700-60	Sequence 60, Appl	336	12	85.7	601	4	US-09-949-016-112334	Sequence 112334,
c 264	12	85.7	53	3	US-09-079-570B-9	Sequence 9, Appl	337	12	85.7	601	4	US-09-949-016-112335	Sequence 112335,
c 265	12	85.7	53	3	US-09-079-570B-10	Sequence 10, Appl	338	12	85.7	601	4	US-09-949-016-112336	Sequence 112336,
c 266	12	85.7	53	4	US-09-042-488B-15	Sequence 15, Appl	339	12	85.7	601	4	US-09-949-016-112337	Sequence 112337,
c 267	12	85.7	53	4	US-09-042-488B-16	Sequence 16, Appl	340	12	85.7	601	4	US-09-949-016-112338	Sequence 112338,
c 268	12	85.7	66	2	US-08-432-871C-66	Sequence 66, Appl	341	12	85.7	601	4	US-09-949-016-112339	Sequence 112339,
c 269	12	85.7	66	3	US-09-270-956-66	Sequence 66, Appl	342	12	85.7	601	4	US-09-949-016-112340	Sequence 112340,
c 270	12	85.7	127	3	US-08-618-100B-32	Sequence 32, Appl	343	12	85.7	601	4	US-09-949-016-112341	Sequence 112341,
c 271	12	85.7	170	4	US-09-513-999C-15859	Sequence 15859, A	344	12	85.7	601	4	US-09-949-016-112342	Sequence 112342,
c 272	12	85.7	266	4	US-09-513-999C-1909	Sequence 1909, Ap	c 345	12	85.7	601	4	US-09-949-016-120050	Sequence 120050,
c 273	12	85.7	284	4	US-09-471-276-444	Sequence 444, App	c 346	12	85.7	601	4	US-09-949-016-120100	Sequence 120100,
c 274	12	85.7	314	4	US-09-513-999C-28367	Sequence 28367, A	c 347	12	85.7	601	4	US-09-949-016-120150	Sequence 120150,
c 275	12	85.7	387	4	US-09-902-540-9183	Sequence 9183, Ap	348	12	85.7	601	4	US-09-949-016-127532	Sequence 127532,
c 276	12	85.7	399	4	US-09-252-991A-14626	Sequence 14626, A	c 349	12	85.7	601	4	US-09-949-016-131274	Sequence 131274,
c 277	12	85.7	468	4	US-09-252-991A-7286	Sequence 7286, Ap	c 350	12	85.7	601	4	US-09-949-016-131725	Sequence 131725,
c 278	12	85.7	570	4	US-09-502-540-5326	Sequence 5326, Ap	c 351	12	85.7	601	4	US-09-949-016-148781	Sequence 148781,
c 279	12	85.7	597	4	US-09-252-991A-7329	Sequence 7329, Ap	c 352	12	85.7	601	4	US-09-949-016-148782	Sequence 148782,
c 280	12	85.7	601	4	US-09-949-016-18580	Sequence 18580, A	353	12	85.7	601	4	US-09-949-016-148783	Sequence 148783,
c 281	12	85.7	601	4	US-09-949-016-18581	Sequence 18581, A	354	12	85.7	601	4	US-09-949-016-148784	Sequence 148784,
c 282	12	85.7	601	4	US-09-949-016-26331	Sequence 26331, A	355	12	85.7	601	4	US-09-949-016-148785	Sequence 148785,
c 283	12	85.7	601	4	US-09-949-016-27900	Sequence 27900, A	c 356	12	85.7	601	4	US-09-949-016-148810	Sequence 148810,
c 284	12	85.7	601	4	US-09-949-016-33270	Sequence 33270, A	c 357	12	85.7	601	4	US-09-949-016-155604	Sequence 155604,
c 285	12	85.7	601	4	US-09-949-016-33271	Sequence 33271, A	c 358	12	85.7	601	4	US-09-949-016-155605	Sequence 155605,
c 286	12	85.7	601	4	US-09-949-016-33272	Sequence 33272, A	c 359	12	85.7	601	4	US-09-949-016-155606	Sequence 155606,
c 287	12	85.7	601	4	US-09-949-016-33276	Sequence 33276, A	c 360	12	85.7	601	4	US-09-949-016-155607	Sequence 155607,
c 288	12	85.7	601	4	US-09-949-016-37007	Sequence 37007, A	c 361	12	85.7	601	4	US-09-949-016-155608	Sequence 155608,
c 289	12	85.7	601	4	US-09-949-016-37008	Sequence 37008, A	c 362	12	85.7	601	4	US-09-949-016-155703	Sequence 155703,
c 290	12	85.7	601	4	US-09-949-016-37009	Sequence 37009, A	c 363	12	85.7	601	4	US-09-949-016-155704	Sequence 155704,
c 291	12	85.7	601	4	US-09-949-016-37010	Sequence 37010, A	c 364	12	85.7	601	4	US-09-949-016-155705	Sequence 155705,
c 292	12	85.7	601	4	US-09-949-016-39549	Sequence 39549, A	c 365	12	85.7	601	4	US-09-949-016-155706	Sequence 155706,
c 293	12	85.7	601	4	US-09-949-016-39550	Sequence 39550, A	c 366	12	85.7	601	4	US-09-949-016-155707	Sequence 155707,
c 294	12	85.7	601	4	US-09-949-016-39582	Sequence 39582, A	c 367	12	85.7	601	4	US-09-949-016-160754	Sequence 160754,
c 295	12	85.7	601	4	US-09-949-016-39583	Sequence 39583, A	c 368	12	85.7	601	4	US-09-949-016-162430	Sequence 162430,
c 296	12	85.7	601	4	US-09-949-016-39644	Sequence 39644, A	c 369	12	85.7	601	4	US-09-949-016-164994	Sequence 164994,
c 297	12	85.7	601	4	US-09-949-016-39645	Sequence 39645, A	c 370	12	85.7	601	4	US-09-949-016-167930	Sequence 167930,
c 298	12	85.7	601	4	US-09-949-016-41246	Sequence 41246, A	371	12	85.7	601	4	US-09-949-016-177931	Sequence 177931,
c 299	12	85.7	601	4	US-09-949-016-42825	Sequence 42825, A	372	12	85.7	601	4	US-09-949-016-177937	Sequence 177937,
c 300	12	85.7	601	4	US-09-949-016-47471	Sequence 47471, A	373	12	85.7	601	4	US-09-949-016-179318	Sequence 179318,
c 301	12	85.7	601	4	US-09-949-016-47472	Sequence 47472, A	c 374	12	85.7	601	4	US-09-949-016-178811	Sequence 178811,
c 302	12	85.7	601	4	US-09-949-016-47473	Sequence 47473, A	c 375	12	85.7	601	4	US-09-949-016-178812	Sequence 178812,
c 303	12	85.7	601	4	US-09-949-016-50103	Sequence 50103, A	376	12	85.7	601	4	US-09-949-016-189945	Sequence 189945,
c 304	12	85.7	601	4	US-09-949-016-62238	Sequence 62238, A	c 377	12	85.7	601	4	US-09-949-016-193771	Sequence 193771,
c 305	12	85.7	601	4	US-09-949-016-62239	Sequence 62239, A	c 378	12	85.7	601	4	US-09-949-016-193772	Sequence 193772,
c 306	12	85.7	601	4	US-09-949-016-62309	Sequence 62309, A	c 379	12	85.7	601	4	US-09-949-016-195485	Sequence 195485,
c 307	12	85.7	601	4	US-09-949-016-62310	Sequence 62310, A	380	12	85.7	601	4	US-09-949-016-195486	Sequence 195486,
c 308	12	85.7	601	4	US-09-949-016-62380	Sequence 62380, A	c 381	12	85.7	601	4	US-09-949-016-195786	Sequence 195786,
c 309	12	85.7	601	4	US-09-949-016-62381	Sequence 62381, A	c 382	12	85.7	601	4	US-09-949-016-196810	Sequence 196810,
c 310	12	85.7	601	4	US-09-949-016-62451	Sequence 62451, A	c 383	12	85.7	601	4	US-09-949-016-196838	Sequence 196838,
c 311	12	85.7	601	4	US-09-949-016-62452	Sequence 62452, A	c 384	12	85.7	669	4	US-09-902-540-5320	Sequence 5320, Ap
c 312	12	85.7	601	4	US-09-949-016-67148	Sequence 67148, A	c 385	12	85.7	715	3	US-08-998-416-709	Sequence 709, App
c 313	12	85.7	601	4	US-09-949-016-67149	Sequence 67149, A	c 386	12	85.7	731	4	US-09-976-594-191	Sequence 191, App
c 314	12	85.7	601	4	US-09-949-016-67139	Sequence 70139, A	c 387	12	85.7	762	4	US-09-902-540-7988	Sequence 7988, Ap
c 315	12	85.7	601	4	US-09-949-016-70139	Sequence 70139, A	c 388	12	85.7	825	4	US-09-821-167-4	Sequence 4, Appl1
c 316	12	85.7	601	4	US-09-949-016-102897	Sequence 102897, A	c 389	12	85.7	838	4	US-09-976-594-56	Sequence 56, Appl
c 317	12	85.7	601	4	US-09-949-016-102898	Sequence 102898, A	c 390	12	85.7	956	3	US-08-755-587-17	Sequence 17, Appl
c 318	12	85.7	601	4	US-09-949-016-104233	Sequence 104233, A	c 391	12	85.7	1008	4	US-09-902-587-17	Sequence 17, Appl
c 319	12	85.7	601	4	US-09-949-016-104234	Sequence 104234, A	c 392	12	85.7	1088	3	US-09-040-285A-1	Sequence 1, Appl1
					Sequence 104304,								

393 12 85.7 1162 2 US-08-726-306A-52 Sequence 52, Appl
c 394 12 85.7 1166 4 US-09-270-767-13064 Sequence 13064, A
395 12 85.7 1192 4 US-09-774-528-90 Sequence 90, Appl
c 396 12 85.7 1230 4 US-09-252-991A-7646 Sequence 7646, Ap
c 397 12 85.7 1272 4 US-09-328-352-1541 Sequence 1541, Ap
398 12 85.7 1278 4 US-09-573-080A-187 Sequence 187, App
399 12 85.7 1334 3 US-09-347-798-11 Sequence 11, Appl
c 400 12 85.7 1395 4 US-09-902-540-3297 Sequence 3297, Ap
c 401 12 85.7 1409 4 US-09-270-767-13790 Sequence 13790, A
c 402 12 85.7 1416 4 US-09-134-000C-2823 Sequence 2823, Ap
403 12 85.7 1496 4 US-09-205-258-214 Sequence 214, App
404 12 85.7 1581 4 US-09-949-016-243 Sequence 243, App
405 12 85.7 1588 4 US-09-949-016-1165 Sequence 1165, Ap
406 12 85.7 1746 4 US-09-573-080A-186 Sequence 186, App
407 12 85.7 1815 4 US-09-799-451-928 Sequence 928, App
408 12 85.7 1958 3 US-09-401-476-1 Sequence 1, Appl
409 12 85.7 2067 2 US-08-713-928B-8 Sequence 8, Appl
c 410 12 85.7 2095 3 US-09-401-476-3 Sequence 3, Appl
c 411 12 85.7 2109 4 US-09-902-540-8825 Sequence 8825, Ap
412 12 85.7 2155 3 US-09-191-171-4 Sequence 4, Appl
413 12 85.7 2155 3 US-09-385-707-4 Sequence 4, Appl
414 12 85.7 2155 4 US-09-639-696C-4 Sequence 30, Appl
415 12 85.7 2155 4 US-09-917-254-30 Sequence 170, App
416 12 85.7 2184 3 US-08-755-587-170 Sequence 111, App
417 12 85.7 2249 4 US-09-205-258-111 Sequence 211, App
418 12 85.7 2335 4 US-09-774-528-211 Sequence 1, Appl
419 12 85.7 2453 5 PCT-US95-07180-1 Sequence 46, Appl
420 12 85.7 2484 3 US-09-276-531-46 Sequence 424, App
421 12 85.7 2585 4 US-09-902-540-424 Sequence 607, App
c 422 12 85.7 2642 4 US-09-949-016-1282 Sequence 1282, Ap
c 423 12 85.7 2737 4 US-09-949-016-607 Sequence 30, Appl
c 424 12 85.7 2790 4 US-09-904-615-30 Sequence 291, App
c 425 12 85.7 2799 4 US-09-495-050A-291 Sequence 487, App
c 426 12 85.7 2896 4 US-09-902-540-487 Sequence 4, Appl
c 427 12 85.7 3128 4 US-10-087-402-4 Sequence 5200, Ap
c 428 12 85.7 3251 4 US-09-949-016-5200 Sequence 813, App
c 429 12 85.7 3252 4 US-09-949-016-813 Sequence 4355, Ap
c 430 12 85.7 3269 4 US-09-949-016-4355 Sequence 4356, Ap
c 431 12 85.7 3269 4 US-09-949-016-4356 Sequence 4733, Ap
c 432 12 85.7 3333 4 US-09-902-540-4753 Sequence 1, Appl
c 433 12 85.7 3566 2 US-08-415-788-1 Sequence 4, Appl
c 434 12 85.7 3566 2 US-08-415-788-4 Sequence 6, Appl
c 435 12 85.7 3566 2 US-08-415-788-6 Sequence 3, Appl
c 436 12 85.7 3567 3 US-08-637-732A-3 Sequence 70, Appl
c 437 12 85.7 4043 4 US-09-814-915A-70 Sequence 1, Appl
c 438 12 85.7 4047 4 US-09-081-385-1 Sequence 147, App
c 439 12 85.7 4047 4 US-09-949-016-977 Sequence 977, App
c 440 12 85.7 4049 4 US-09-081-385-147 Sequence 7, Appl
c 441 12 85.7 4480 3 US-09-191-171-7 Sequence 7, Appl
c 442 12 85.7 4480 3 US-09-385-707-7 Sequence 790, App
c 443 12 85.7 4793 4 US-09-902-540-790 Sequence 810, App
c 444 12 85.7 5337 4 US-09-902-540-810 Sequence 7, Appl
c 445 12 85.7 5375 3 US-08-757-223-7 Sequence 7, Appl
c 446 12 85.7 5481 4 US-09-949-016-12049 Sequence 12049, A
c 447 12 85.7 5484 4 US-09-949-016-15589 Sequence 15589, A
c 448 12 85.7 5484 4 US-09-949-016-15589 Sequence 13664, A
c 449 12 85.7 5695 4 US-09-949-016-13664 Sequence 11, Appl
c 450 12 85.7 5847 4 US-09-962-665-11 Sequence 11, Appl
451 12 85.7 5847 4 US-09-963-333-11 Sequence 11, Appl
452 12 85.7 5847 4 US-09-962-677-11 Sequence 11, Appl
453 12 85.7 6021 4 US-09-949-016-13178 Sequence 13178, A
454 12 85.7 6021 4 US-09-949-016-13178 Sequence 13178, A
c 455 12 85.7 6200 3 US-09-439-923-1 Sequence 1, Appl
456 12 85.7 6200 4 US-09-711-202A-1 Sequence 1, Appl
457 12 85.7 6200 4 US-09-711-205A-1 Sequence 6, Appl
458 12 85.7 6238 4 US-09-639-696C-6 Sequence 15446, A
459 12 85.7 6289 4 US-09-949-016-15446 Sequence 12588, A
c 460 12 85.7 6299 4 US-09-949-016-12588 Sequence 12920, A
461 12 85.7 6501 4 US-09-949-016-12920 Sequence 13771, A
462 12 85.7 6644 4 US-09-949-016-13771 Sequence 13771, A
463 12 85.7 6711 4 US-09-949-016-16879 Sequence 16879, A
464 12 85.7 6752 4 US-09-949-016-14776 Sequence 14776, A
465 12 85.7 7626 4 US-09-949-016-14216 Sequence 14216, A

ALIGNMENTS

RESULT 1
US-08-472-217-1/c
; Sequence 1, Application US/08472217
; Patent No. 5726058
; GENERAL INFORMATION:
; APPLICANT: Alanen-Kurki, Leena
; APPLICANT: Auvinen, Petri
; APPLICANT: Jaakkola, Panu
; APPLICANT: Jalkanen, Markku
; APPLICANT: Lepp, Sirpa
; APPLICANT: Mali, Markku
; APPLICANT: Vihtinen, Tapani
; APPLICANT: W rri, Anni
; TITLE OF INVENTION: Syndecan Stimulation Of Cellular
; TITLE OF INVENTION: Differentiation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,217
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/206,186
FILING DATE: 07-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,427
FILING DATE: 01-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 1102.0050003
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 26700 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: Join(4378..4443, 22026..22106, 23001..23483,
23905..24039, 24251..24418)
US-08-472-217-1

Query Match 100.0%; Score 14; DB 1; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
DB 4027 TTGCCCCGCCCTT 4014

RESULT 2
US-08-488-199-5/c
Sequence 5, Application US/08488199
Patent No. 5851993
GENERAL INFORMATION:
APPLICANT: Jalkanen, Markku
APPLICANT: Mali, Markku
TITLE OF INVENTION: Suppression of Tumor Cell Growth By
TITLE OF INVENTION: Syndecan-1 Ectodomain
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,199
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,862
FILING DATE: 13-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 1102.0130001
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 26700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 4378..4443
FEATURE:
NAME/KEY: CDS
LOCATION: 22026..22107
FEATURE:
NAME/KEY: CDS
LOCATION: 23002..23483
FEATURE:
NAME/KEY: CDS
LOCATION: 23905..24040
FEATURE:
NAME/KEY: CDS
LOCATION: 24252..24418
US-08-488-199-5

Query Match 100.0%; Score 14; DB 2; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
DB 4027 TTGCCCCGCCCTT 4014

RESULT 3
US-08-760-534A-1/c
Sequence 1, Application US/08760534A
Patent No. 6017727
GENERAL INFORMATION:
APPLICANT: JALKANEN, MARKKU
APPLICANT: JAAKKOLA, PANU
APPLICANT: VIHINEN, TAPANI
TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN
TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,534A
FILING DATE: 02-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,186
FILING DATE: 07-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FI93/00514
FILING DATE: 01-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: CIMBALA, MICHELE A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

; LENGTH: 26700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4378..4443, 22026..22106, 23001..23483,
; LOCATION: 23905..24039, 24251..24418)
US-08-760-534A-1

Query Match 100.0%; Score 14; DB 3; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||
DB 4027 TTGCCCCGCCCTT 4014

RESULT 4
US-09-336-757-1/C
; Sequence 1, Application US/09336757
; Patent No. 6492344
; GENERAL INFORMATION:
; APPLICANT: JAAKANEN, MARKKU
; APPLICANT: JAAKKOLA, PANU
; APPLICANT: VIHINEN, TAPANI
; TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN
; TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/336,757
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/760,534
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI93/00514
; FILING DATE: 01-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CIMBALA, MICHELE A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4378..4443, 22026..22106, 23001..23483,
; LOCATION: 23905..24039, 24251..24418)
US-09-336-757-1

Query Match 100.0%; Score 14; DB 4; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||
DB 4027 TTGCCCCGCCCTT 4014

RESULT 5
US-09-270-767-6481
; Sequence 6481, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6481
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6481

Query Match 92.9%; Score 13; DB 4; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 379 TGCCCCGCCCTT 391

RESULT 6
US-09-270-767-21763
; Sequence 21763, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21763
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-21763

Query Match 92.9%; Score 13; DB 4; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 379 TGCCCCGCCCTT 391

RESULT 7
US-09-949-016-21120
; Sequence 21120, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21120
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21120

Query Match 92.9%; Score 13; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
DB 240 TTGCCCCGCCCTT 252

RESULT 8

US-09-949-016-47488
; Sequence 47488, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47488
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-47488

Query Match 92.9%; Score 13; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
DB 240 TTGCCCCGCCCTT 252

RESULT 9

US-09-949-016-50128/c
; Sequence 50128, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50128
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-50128

Query Match 92.9%; Score 13; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
DB 62 TGCCCCGCCCTT 50

RESULT 10

US-09-949-016-50129/c
; Sequence 50129, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50129
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-50129

Query Match 92.9%; Score 13; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
DB 54 TGCCCCGCCCTT 42

RESULT 11

US-09-949-016-94941/c
; Sequence 94941, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94941

;
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-94941

Query Match 92.9%; Score 13; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
Db 591 TTGCCCCGCCCTT 579

RESULT 12
US-09-621-976-1968
; Sequence 1968, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1968
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 142..411
; NAME/KEY: sig_peptide
; LOCATION: 142..366
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 3.70000004768372
; OTHER INFORMATION: seq LLLRLQSTALQIQ/MQ
US-09-621-976-1968

Query Match 92.9%; Score 13; DB 4; Length 609;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
Db 194 TGCCCCGCCCTT 206

RESULT 13
US-08-998-416-459
; Sequence 459, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 459:
SEQUENCE CHARACTERISTICS:
LENGTH: 794 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1330UP
US-08-998-416-459

Query Match 92.9%; Score 13; DB 3; Length 794;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
Db 525 TGCCCCGCCCTT 537

RESULT 14
US-09-252-991A-3585/c
; Sequence 3585, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3585
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3585

Query Match 92.9%; Score 13; DB 4; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
Db 123 TTGCCCCGCCCTT 111

RESULT 15
US-09-522-714-7
; Sequence 7, Application US/09522714


```
; Patent No. 6563020
; GENERAL INFORMATION:
; APPLICANT: Simmons, Carl R.
; APPLICANT: Yalpani, Nasser
; TITLE OF INVENTION: Maize Chitinases and Their Use in
; TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants
; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522,714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125,915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1195
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (41)...(1009)
US-09-522-714-7

Query Match          92.9%; Score 13; DB 4; Length 1195;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
Db      450 TGCCCCGCCCCCTT 462

RESULT 16
US-09-620-312D-867
; Sequence 867, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 867_
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (307)...(1068)
US-09-620-312D-867

Query Match          92.9%; Score 13; DB 4; Length 1326;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
Db      185 TGCCCCGCCCCCTT 197

RESULT 17
5242798-6/c
; Patent No. 5242798
; APPLICANT: SUTCLIFFE, J. GERGOR
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES CORRESPONDING
; TO PORTIONS OF PROTEINOLIDS TRANSLATED FROM BRAIN-SPECIFIC MRNAS,
; RECEPTORS, METHODS AND DIAGNOSTICS USING THE SAME
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/476,961
; FILING DATE: 07-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 58,620
; FILING DATE: 03-JUN-1987
; APPLICATION NUMBER: 516,136
; FILING DATE: 21-JUL-1983
; SEQ ID NO:6:
; LENGTH: 1409
5242798-6

Query Match          92.9%; Score 13; DB 6; Length 1409;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
Db      1095 TGCCCCGCCCCCTT 1083

RESULT 18
5242798-6/c
; Patent No. 5242798
; APPLICANT: SUTCLIFFE, J. GERGOR
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES CORRESPONDING
; TO PORTIONS OF PROTEINOLIDS TRANSLATED FROM BRAIN-SPECIFIC MRNAS,
; RECEPTORS, METHODS AND DIAGNOSTICS USING THE SAME
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/476,961
; FILING DATE: 07-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 58,620
; FILING DATE: 03-JUN-1987
; APPLICATION NUMBER: 516,136
; FILING DATE: 21-JUL-1983
; SEQ ID NO:6:
; LENGTH: 1409
5242798-6

Query Match          92.9%; Score 13; DB 6; Length 1409;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
Db      1095 TGCCCCGCCCCCTT 1083

RESULT 19
US-09-149-476-297
; Sequence 297, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 92.9%; Score 13; DB 3; Length 1910;

Best Local Similarity 100.0%; Pred. No. 1.3e+03; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCGCCCTT 14

Db 1292 TGCCCGCCCTT 1304

RESULT 20

US-09-149-476-164
Sequence 164, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162

EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11

EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match Similarity 92.9%; Score 13; DB 3; Length 1945;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

|||||

Db 1314 TGCCCCGCCCTT 1326

RESULT 21

US-09-489-039A-5810/c
Sequence 5810, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 5810
LENGTH: 2058
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5810

```
Query Match          92.9%; Score 13; DB 4; Length 2058;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
DB      413 TGCCCCGCCCCCTT 401

RESULT 22
US-09-949-016-17058
; Sequence 17058, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17058
; LENGTH: 4812
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17058

Query Match          92.9%; Score 13; DB 4; Length 4812;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
DB      232 TGCCCCGCCCCCTT 244

RESULT 23
US-09-902-540-907
; Sequence 907, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 907
; LENGTH: 6821
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(6821)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-907

Query Match          92.9%; Score 13; DB 4; Length 6821;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCCCTT 14
DB      3442 TGCCCCGCCCCCTT 3454

RESULT 24
US-09-949-016-14510
; Sequence 14510, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14510
; LENGTH: 7846
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(7846)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14510

Query Match          92.9%; Score 13; DB 4; Length 7846;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCCCT 13
DB      2806 TTGCCCCGCCCCCT 2818

RESULT 25
US-09-949-016-14077/c
; Sequence 14077, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14077
; LENGTH: 8641
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(8641)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14077

Query Match          92.9%; Score 13; DB 4; Length 8641;
```

Best Local Similarity 100.0%; Pred. No. 1.3e+03; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

Db 1213 TGCCCCGCCCTT 1201

RESULT 26

US-09-949-016-13551
; Sequence 13551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13551
; LENGTH: 9620
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13551

Query Match 92.9%; Score 13; DB 4; Length 9620;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

Db 4270 TGCCCCGCCCTT 4282

RESULT 27

US-09-949-016-14030
; Sequence 14030, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14030
; LENGTH: 11231
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14030

Query Match 92.9%; Score 13; DB 4; Length 11231;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

|||||

Db 1188 TGCCCCGCCCTT 1200

RESULT 28

US-09-949-016-16030
; Sequence 16030, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16030
; LENGTH: 12163
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16030

Query Match 92.9%; Score 13; DB 4; Length 12163;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

|||||

Db 1502 TGCCCCGCCCTT 1514

RESULT 29

US-09-949-016-12202
; Sequence 12202, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12202
; LENGTH: 12164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12202

Query Match 92.9%; Score 13; DB 4; Length 12164;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14

|||||

Db 1502 TGCCCCGCCCTT 1514

RESULT 30

US-09-949-016-11927

; Sequence 11927, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11927
; LENGTH: 15585
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(15585)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11927

Query Match 92.9%; Score 13; DB 4; Length 15585;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14
Db 1902 TGCCCCGCCCTT 1914

RESULT 31
US-09-949-016-15627
; Sequence 15627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15627
; LENGTH: 15585
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(15585)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15627

Query Match 92.9%; Score 13; DB 4; Length 15585;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14
Db 1902 TGCCCCGCCCTT 1914

RESULT 32
US-09-949-016-13134/c
; Sequence 13134, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13134
; LENGTH: 31467
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(31467)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13134

Query Match 92.9%; Score 13; DB 4; Length 31467;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCT 13
Db 17125 TTGCCCCGCCCT 17113

RESULT 33
US-09-949-016-11907/c
; Sequence 11907, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11907
; LENGTH: 31868
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(31868)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11907

Query Match 92.9%; Score 13; DB 4; Length 31868;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCT 13

```
Db 17526 TTGCCCCGCCCTT 17514
|||||
RESULT 34
US-09-949-016-14283
; Sequence 14283, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14283
; LENGTH: 33748
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(33748)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14283

Query Match 92.9%; Score 13; DB 4; Length 33748;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||||
Db 547 TGCCCCGCCCTT 559
|||||
RESULT 35
US-09-902-540-1266/c
; Sequence 1266, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1266
; LENGTH: 41768
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(41768)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1266

Query Match 92.9%; Score 13; DB 4; Length 41768;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||||
Db 17526 TTGCCCCGCCCTT 17514
|||||
RESULT 36
US-09-949-016-12572/c
; Sequence 12572, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12572
; LENGTH: 53806
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(53806)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12572

Query Match 92.9%; Score 13; DB 4; Length 53806;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
|||||
Db 39788 TTGCCCCGCCCTT 39776
|||||
RESULT 37
US-09-949-016-14209/c
; Sequence 14209, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14209
; LENGTH: 83210
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(83210)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14209

Query Match 92.9%; Score 13; DB 4; Length 83210;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
|||||
```


Db 1854 TTGCCCCGCCCTT 1842

RESULT 38

US-09-949-016-13208/c
; Sequence 13208, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13208
; LENGTH: 97989
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(97989)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13208

Query Match 92.9%; Score 13; DB 4; Length 97989;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14

|||||

Db 17072 TGCCCCGCCCTT 17060

RESULT 39

US-09-949-016-16933
; Sequence 16933, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16933
; LENGTH: 101011
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16933

Query Match 92.9%; Score 13; DB 4; Length 101011;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14

|||||

Db 1873 TGCCCCGCCCTT 1885

RESULT 40

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 92.9%; Score 13; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14

|||||

Db 4154353 TGCCCCGCCCTT 4154365

RESULT 41

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 92.9%; Score 13; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14

|||||

Db 4162102 TGCCCCGCCCTT 4162114

RESULT 42

US-08-556-978B-34
; Sequence 34, Application US/08556978B
; Patent No. 6268169

GENERAL INFORMATION:
APPLICANT: FAHNESTOCK, STEPHEN F.
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
TITLE OF INVENTION: SPIDER SILK ANALOGS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556,978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-34

Query Match 88.6%; Score 12.4; DB 3; Length 39;
Best Local Similarity 92.9%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
DB 26 TTGCCCCGCCCTT 39
RESULT 43
US-08-343-443B-46/c
Sequence 46, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougaastel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLOCATIONS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA

COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-343-443B-46

Query Match 88.6%; Score 12.4; DB 2; Length 69;
Best Local Similarity 92.9%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
DB 55 TTGCCCCGCCCTT 42

RESULT 44
US-09-626-002-28
Sequence 28, Application US/09626002
Patent No. 6630324
GENERAL INFORMATION:
APPLICANT: Barski, Oleg
APPLICANT: Aguilar-Cordova, Estuardo
APPLICANT: Bohren, Kurt
APPLICANT: Gabbay, Kenneth
TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
FILE REFERENCE: P01859US1/09903803/OTA 99-53
CURRENT APPLICATION NUMBER: US/09/626,002
CURRENT FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: US 60/146,266
PRIOR FILING DATE: 1999-07-29
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn version 3.0
SEQ ID NO 28
LENGTH: 94
TYPE: DNA
ORGANISM: Human
US-09-626-002-28

Query Match 88.6%; Score 12.4; DB 4; Length 94;
Best Local Similarity 92.9%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
DB 35 TTGCCCCGCCCTT 48

```
RESULT 45
US-09-626-002-27
; Sequence 27, Application US/09626002
; Patent No. 6630324
; GENERAL INFORMATION:
; APPLICANT: Barski, Oleg
; APPLICANT: Aguilar-Cordova, Estuardo
; APPLICANT: Bohren, Kurt
; APPLICANT: Gabbay, Kenneth
; TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
; FILE REFERENCE: P01859US1/09903803/OTA 99-53
; CURRENT APPLICATION NUMBER: US/09/626,002
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: US 60/146,266
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 95
; TYPE: DNA
; ORGANISM: Human
US-09-626-002-27

Query Match      88.6%; Score 12.4; DB 4; Length 95;
Best Local Similarity 92.9%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
Db 36 TGGCCCCGCCCCCTT 49

RESULT 46
US-09-626-002-26
; Sequence 26, Application US/09626002
; Patent No. 6630324
; GENERAL INFORMATION:
; APPLICANT: Barski, Oleg
; APPLICANT: Aguilar-Cordova, Estuardo
; APPLICANT: Bohren, Kurt
; APPLICANT: Gabbay, Kenneth
; TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
; FILE REFERENCE: P01859US1/09903803/OTA 99-53
; CURRENT APPLICATION NUMBER: US/09/626,002
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: US 60/146,266
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 113
; TYPE: DNA
; ORGANISM: Human
US-09-626-002-26

Query Match      88.6%; Score 12.4; DB 4; Length 113;
Best Local Similarity 92.9%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
Db 54 TGGCCCCGCCCCCTT 67

RESULT 47
US-09-626-002-25
; Sequence 25, Application US/09626002
; Patent No. 6630324
; GENERAL INFORMATION:
; APPLICANT: Barski, Oleg
; APPLICANT: Aguilar-Cordova, Estuardo
; APPLICANT: Bohren, Kurt
; APPLICANT: Gabbay, Kenneth
; TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
; FILE REFERENCE: P01859US1/09903803/OTA 99-53
; CURRENT APPLICATION NUMBER: US/09/626,002
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: US 60/146,266
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 137
; TYPE: DNA
; ORGANISM: Human
US-09-626-002-25

Query Match      88.6%; Score 12.4; DB 4; Length 137;
Best Local Similarity 92.9%; Pred. No. 2.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
Db 78 TGGCCCCGCCCCCTT 91

RESULT 48
US-09-626-002-24
; Sequence 24, Application US/09626002
; Patent No. 6630324
; GENERAL INFORMATION:
; APPLICANT: Barski, Oleg
; APPLICANT: Aguilar-Cordova, Estuardo
; APPLICANT: Bohren, Kurt
; APPLICANT: Gabbay, Kenneth
; TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
; FILE REFERENCE: P01859US1/09903803/OTA 99-53
; CURRENT APPLICATION NUMBER: US/09/626,002
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: US 60/146,266
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 153
; TYPE: DNA
; ORGANISM: Human
US-09-626-002-24

Query Match      88.6%; Score 12.4; DB 4; Length 153;
Best Local Similarity 92.9%; Pred. No. 2.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
Db 94 TGGCCCCGCCCCCTT 107

RESULT 49
US-09-626-002-30/c
; Sequence 30, Application US/09626002
; Patent No. 6630324
; GENERAL INFORMATION:
; APPLICANT: Barski, Oleg
; APPLICANT: Aguilar-Cordova, Estuardo
; APPLICANT: Bohren, Kurt
; APPLICANT: Gabbay, Kenneth
; TITLE OF INVENTION: Aldehyde Reductase Bidirectional Promoter and Its Use
; FILE REFERENCE: P01859US1/09903803/OTA 99-53
; CURRENT APPLICATION NUMBER: US/09/626,002
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: US 60/146,266
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
```

; LENGTH: 153
; TYPE: DNA
; ORGANISM: Human
US-09-626-002-30

Query Match 88.6%; Score 12.4; DB 4; Length 153;
Best Local Similarity 92.9%; Pred. No. 2.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
Db 60 TGGCCCCGCCCTT 47

RESULT 50
US-09-513-999C-14750/c
; Sequence 14750, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14750
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14750

Query Match 88.6%; Score 12.4; DB 4; Length 202;
Best Local Similarity 92.9%; Pred. No. 2.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
Db 39 TGGCCCCGCCCTT 26

Search completed: April 25, 2005, 15:11:08
Job time : 70 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:06:08 ; Search time 139.638 Seconds
(without alignments)
593.509 Million cell updates/sec

Title: US-10-010-476-15

Perfect score: 14
Sequence: 1 TTGCCCCGCCCTT 14

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 500 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	2	Aax40730 GC-box e-
2	14	100.0	22	2	Aax40729 GC-box d-
3	14	100.0	30	2	Aax40725 GC-box b-
c 4	14	100.0	30	2	Aax40724 GC-box a
5	14	100.0	50	2	Aax40728 GC-box c-
c 6	14	100.0	255	10	Ac98140 Human col
7	14	100.0	292	8	Abx35297 Bovine ES
c 8	14	100.0	326	4	Aas39325 Novel hum
c 9	14	100.0	340	4	Aas39329 Novel hum
c 10	14	100.0	411	5	Aaf65653 Novel hum
c 11	14	100.0	587	5	Aas34163 Human cDN
12	14	100.0	795	4	Aaf84400 Human oli
c 13	14	100.0	1062	4	Abi04361 Drosophil
c 14	14	100.0	1359	5	Aas65178 DNA encod
c 15	14	100.0	1563	2	Aaz53007 Human pro
c 16	14	100.0	1597	5	Aas65180 DNA encod
c 17	14	100.0	1597	13	Acn40708 Tumour-as
c 18	14	100.0	1597	13	Adf66182 Human pro
c 19	14	100.0	1619	12	Adq84817 Human tum
c 20	14	100.0	1619	12	Adq85892 Human tum

c 21	14	100.0	1622	4	AAH14860
c 22	14	100.0	1655	13	ADR21626
c 23	14	100.0	2064	5	AAS79453
24	14	100.0	2573	10	ABT42316
25	14	100.0	3170	6	ABA05737
26	14	100.0	3250	6	ABA05738
27	14	100.0	3665	4	ABL11522
c 28	14	100.0	4226	4	ABL09210
c 29	14	100.0	4878	4	ABL04360
30	14	100.0	8570	12	ADO20320
c 31	14	100.0	8570	13	ADP55596
c 32	14	100.0	9760	8	ABZ74489
c 33	14	100.0	9760	8	ADA44498
34	14	100.0	13273	4	AAS36849
35	14	100.0	13273	10	ADE47543
36	14	100.0	13273	13	ADJ08961
c 37	14	100.0	17131	3	AZ60888
c 38	14	100.0	26598	2	AAU08126
c 39	14	100.0	26700	2	AAU08126
c 40	14	100.0	26700	2	AAV15946
c 41	14	100.0	26700	2	AAV81283
c 42	14	100.0	26700	8	ACA60750
c 43	14	100.0	51001	12	ADJ10262
c 44	13	92.9	27	3	AAH71425
45	13	92.9	34	6	ABK98227
c 46	13	92.9	35	6	ABK98226
c 47	13	92.9	40	6	ABK98129
c 48	13	92.9	207	13	ADS47155
c 49	13	92.9	238	6	AAU43323
50	13	92.9	260	10	ADC21271
51	13	92.9	260	10	ADD68310
52	13	92.9	260	10	ADD29394
53	13	92.9	260	10	ABZ70045
54	13	92.9	265	7	ADS65310
55	13	92.9	333	10	ADC21272
56	13	92.9	333	10	ADD68311
57	13	92.9	333	10	ADD29395
58	13	92.9	333	10	ABZ70046
59	13	92.9	397	2	AAU73660
c 60	13	92.9	424	8	ACH04308
c 61	13	92.9	466	8	ABZ18450
62	13	92.9	467	9	ACH34271
63	13	92.9	483	6	ABN95835
c 64	13	92.9	529	12	ADQ17204
c 65	13	92.9	534	12	ACH79261
c 66	13	92.9	544	10	ADB56762
c 67	13	92.9	544	10	ABT41165
c 68	13	92.9	544	12	ADP72113
c 69	13	92.9	581	4	ABA59739
c 70	13	92.9	581	4	AAI39606
c 71	13	92.9	581	4	AAK33880
c 72	13	92.9	581	4	AAK08008
c 73	13	92.9	581	4	ABS33689
c 74	13	92.9	581	6	ABS08713
c 75	13	92.9	658	3	AAAB1686
76	13	92.9	677	3	AAAG3869
77	13	92.9	679	3	AAH72319
c 78	13	92.9	707	4	AAH29824
c 79	13	92.9	741	11	ADJ12009
80	13	92.9	794	2	ADR01767
c 81	13	92.9	831	11	ABD04981
82	13	92.9	903	12	ADJ44715
83	13	92.9	969	9	ADA49192
84	13	92.9	969	12	ADJ44454
85	13	92.9	975	10	ADH48177
86	13	92.9	1065	13	ADT48724
c 87	13	92.9	1104	5	ADL63282
88	13	92.9	1195	3	AAAG6225
89	13	92.9	1206	13	ACN41242
90	13	92.9	1255	3	AAU76682
91	13	92.9	1257	13	ACN41241
92	13	92.9	1326	4	AAI58975
93	13	92.9	1326	5	ADQ99197

Aah14860	Human cDN
Adr21626	Human enz
Aas79453	DNA encod
Abt42316	Toxicity
Abao5737	Murine ne
Abao5738	Murine ne
Abli1522	Drosophil
Abi09210	Drosophil
Abi04360	Drosophil
Ado20320	Human PRO
Adp55596	Human PRO
Abz74489	Secreted
Ada44498	Human sec
Aas36849	Human car
Ado8961	Human car
Aaz60888	DNA encod
Aat08126	Mouse syn
Aaq7902	Syndecan
Aav15946	Mouse syn
Aav81283	Mouse syn
ACA60750	Mouse cel
Adj10262	Human ger
Aaa71425	Murine pa
Abk98227	Triple he
Abk98226	Triple he
Abk98129	Triple he
Ado47155	Bacterial
Ado43323	Human DTC
Adc21271	Human cel
Add68310	Human ang
Add29394	Antiarter
Abz70045	Oligonucl
Adse65310	Corn seed
Adc21272	Human cal
Add68311	DNA compr
Adz70046	Human cal
Abz73660	ICAM-2 ge
Ach04308	Human cDN
Abz18450	Group I II
Ach34271	Human end
Abn95835	Gene #233
Adn17204	Human sof
Ach79261	Human gen
Adb56762	Toxicity-
Abt41165	Toxicity
Adp72113	Renal Cox
Abas9739	Human foe
Aai39606	Probe #82
Aak33880	Human bon
Aak08008	Human bra
Abas3689	Human liv
Abso8713	Human gen
AAAB1686	N. mening
AAAG3869	cDNA sequ
Aaa72319	Human Rec
Aah29824	S cerevis
Adj12009	Wheat cDN
ADR01767	A. gossyp
ABD04981	Pseudomon
ADJ44715	Plant cDN
ADA49192	Maize gen
ADJ44454	Plant cDN
ADH48177	Rat gene
ADT48724	Bacterial
ADL63282	Human ova
AAAG6225	cDNA enco
ACN41242	Human dia
AAU76682	Human ORF
ACN41241	Human dia
AAI58975	Human pol
ADQ99197	DNA encod

94	13	92.9	1326	9	ADB48957	Adb48957 Novel hum	167	13	92.9	1657	9	ADA96569	Ada96569 Human PRO
95	13	92.9	1368	4	AA160761	Aa160761 Human pol	168	13	92.9	1657	9	ADA81141	Ada81141 Human PRO
96	13	92.9	1395	13	ACN41239	Acn41239 Human dia	169	13	92.9	1657	9	ADA96017	Ada96017 Human PRO
97	13	92.9	1395	13	ACN41240	Acn41240 Human dia	170	13	92.9	1657	9	ADB26326	Adb26326 cDNA enco
c 98	13	92.9	1404	13	ADT48163	Adt48163 Bacterial	171	13	92.9	1657	9	ADB21811	Adb21811 Novel hum
c 99	13	92.9	1409	1	AAN50474	Aan50474 Sequence	172	13	92.9	1657	9	ADA77590	Ada77590 Human PRO
100	13	92.9	1567	13	ADRO7676	Adr07676 Full leng	173	13	92.9	1657	9	ADB18330	Adb18330 cDNA enco
101	13	92.9	1647	5	ABX71282	Abx71282 Human sig	174	13	92.9	1657	9	ADA87013	Ada87013 Novel hum
102	13	92.9	1657	3	AA265017	Aa265017 Membrane-	175	13	92.9	1657	9	ADA88116	Ada88116 Novel hum
103	13	92.9	1657	4	AA521460	Aa521460 Human cDN	176	13	92.9	1657	9	ADA46504	Ada46504 Novel hum
104	13	92.9	1657	5	AAF44163	Aaf44163 Human PRO	177	13	92.9	1657	9	ADB28534	Adb28534 cDNA enco
105	13	92.9	1657	8	ACA64299	Ac64299 Novel hum	178	13	92.9	1657	9	ADB29086	Adb29086 cDNA enco
106	13	92.9	1657	8	ACA03819	Ac03819 cDNA enco	179	13	92.9	1657	9	ACH65413	Ach65413 Human cDN
107	13	92.9	1657	8	ABX89357	Abx89357 DNA enco	180	13	92.9	1657	9	ADA77038	Ada77038 Human PRO
108	13	92.9	1657	8	ABX80758	Abx80758 Human sec	181	13	92.9	1657	9	ADA22321	Ada22321 Human cDN
109	13	92.9	1657	8	ACD44267	Acd44267 cDNA enco	182	13	92.9	1657	9	ADA88668	Ada88668 Novel hum
110	13	92.9	1657	8	ACD42011	Acd42011 Human sec	183	13	92.9	1657	9	ADA97673	Ada97673 Human PRO
111	13	92.9	1657	8	ABX79438	Abx79438 Human sec	184	13	92.9	1657	9	ADB27430	Adb27430 cDNA enco
112	13	92.9	1657	8	ACA93459	Ac93459 Novel hum	185	13	92.9	1657	9	ADB22363	Adb22363 Novel hum
113	13	92.9	1657	8	ABX81141	Abx81141 Novel hum	186	13	92.9	1657	9	ACD39403	Acd39403 Human cDN
114	13	92.9	1657	8	ACA04240	Ac04240 Human cDN	187	13	92.9	1657	9	ADA06487	Ada06487 Human sec
115	13	92.9	1657	8	ACA92957	Ac92957 Novel hum	188	13	92.9	1657	9	ADA39180	Ada39180 Human cDN
116	13	92.9	1657	8	ABX17041	Abx17041 Human PRO	189	13	92.9	1657	9	ADA67054	Ada67054 Human PRO
117	13	92.9	1657	9	ACA67896	Ac67896 Novel hum	190	13	92.9	1657	9	ADB22915	Adb22915 Human PRO
118	13	92.9	1657	9	ACA88345	Ac88345 Human sec	191	13	92.9	1657	9	ADB23688	Adb23688 Human PRO
119	13	92.9	1657	9	ACD81852	Acd81852 cDNA enco	192	13	92.9	1657	9	ADA92410	Ada92410 Novel hum
120	13	92.9	1657	9	ADA45952	Ad45952 Novel hum	193	13	92.9	1657	9	ADB15473	Adb15473 Human PRO
121	13	92.9	1657	9	ADA76383	Ad76383 Human PRO	194	13	92.9	1657	9	ADB38725	Adb38725 Novel hum
122	13	92.9	1657	9	ADA19033	Ad19033 Human PRO	195	13	92.9	1657	9	ADB96206	Adb96206 Human PRO
123	13	92.9	1657	9	ADA61656	Ad61656 Homo sapi	196	13	92.9	1657	9	ADB38173	Adb38173 Novel hum
124	13	92.9	1657	9	ADB19441	Adb19441 Novel hum	197	13	92.9	1657	9	ADB66645	Adb66645 Novel hum
125	13	92.9	1657	9	ADB27982	Adb27982 cDNA enco	198	13	92.9	1657	10	ADB89725	Adb89725 Human PRO
126	13	92.9	1657	9	ADA86461	Ad86461 Novel hum	199	13	92.9	1657	10	ADB90457	Adb90457 Human PRO
127	13	92.9	1657	9	ADB16025	Adb16025 Human PRO	200	13	92.9	1657	10	ADB39558	Adb39558 Novel hum
128	13	92.9	1657	9	ADA37709	Ad37709 Human cDN	201	13	92.9	1657	10	ADB47181	Adb47181 Novel hum
129	13	92.9	1657	9	ADA47811	Ad47811 Human PRO	202	13	92.9	1657	10	ADB86788	Adb86788 Human PRO
130	13	92.9	1657	9	ADA21395	Ad21395 Human cDN	203	13	92.9	1657	10	ADB77393	Adb77393 Novel hum
131	13	92.9	1657	9	ADA10182	Ad10182 Human cDN	204	13	92.9	1657	10	ADB34550	Adb34550 Human PRO
132	13	92.9	1657	9	ADA67606	Ad67606 Human PRO	205	13	92.9	1657	10	ADB35654	Adb35654 Human PRO
133	13	92.9	1657	9	ADB30613	Adb30613 cDNA enco	206	13	92.9	1657	10	ADB33998	Adb33998 Human PRO
134	13	92.9	1657	9	ADA85909	Ad85909 Novel hum	207	13	92.9	1657	10	ADB35102	Adb35102 Human PRO
135	13	92.9	1657	9	ADA81726	Ad81726 cDNA enco	208	13	92.9	1657	10	ADB36206	Adb36206 Human PRO
136	13	92.9	1657	9	ADA97121	Ad97121 Human PRO	209	13	92.9	1657	10	ADB46601	Adb46601 Novel hum
137	13	92.9	1657	9	ADA79425	Ad79425 Human PRO	210	13	92.9	1657	10	ADC57678	Adc57678 Human PRO
138	13	92.9	1657	9	ADA87564	Ad87564 Novel hum	211	13	92.9	1657	10	ADC55042	Adc55042 Human PRO
139	13	92.9	1657	9	ADB16766	Adb16766 Human PRO	212	13	92.9	1657	10	ADC11909	Adc11909 Human cDN
140	13	92.9	1657	9	ADA27834	Ad27834 Human cDN	213	13	92.9	1657	10	ADC56331	Adc56331 Human PRO
141	13	92.9	1657	9	ADA91858	Ad91858 Novel hum	214	13	92.9	1657	10	ADC07386	Adc07386 Human cDN
142	13	92.9	1657	9	ADB14921	Adb14921 Human PRO	215	13	92.9	1657	10	ADC11376	Adc11376 Human cDN
143	13	92.9	1657	9	ADB18982	Adb18982 Novel hum	216	13	92.9	1657	10	ADC50474	Adc50474 Novel hum
144	13	92.9	1657	9	ADA94097	Ad94097 Human PRO	217	13	92.9	1657	10	ADC72021	Adc72021 Novel hum
145	13	92.9	1657	9	ADB19993	Adb19993 Novel hum	218	13	92.9	1657	10	ADC60000	Adc60000 Novel hum
146	13	92.9	1657	9	ADB13305	Adb13305 Human PRO	219	13	92.9	1657	10	ADC53007	Adc53007 Novel hum
147	13	92.9	1657	9	ACD98640	Ac98640 Novel hum	220	13	92.9	1657	10	ADC57361	Adc57361 Novel hum
148	13	92.9	1657	9	ADA94414	Ad94414 Human cDN	221	13	92.9	1657	10	ADC60552	Adc60552 Novel hum
149	13	92.9	1657	9	ADA74559	Ad74559 Human PRO	222	13	92.9	1657	10	ADC51027	Adc51027 Novel hum
150	13	92.9	1657	9	ADB24792	Adb24792 Human PRO	223	13	92.9	1657	10	ADC65554	Adc65554 Human PRO
151	13	92.9	1657	9	ADA82316	Ad82316 Human PRO	224	13	92.9	1657	10	ADC54652	Adc54652 Novel hum
152	13	92.9	1657	9	ADA75279	Ad75279 Human PRO	225	13	92.9	1657	10	ADC53613	Adc53613 Novel hum
153	13	92.9	1657	9	ADA85357	Ad85357 Novel hum	226	13	92.9	1657	10	ADC59136	Adc59136 Novel hum
154	13	92.9	1657	9	ADA84805	Ad84805 Novel hum	227	13	92.9	1657	10	ADC56014	Adc56014 Novel hum
155	13	92.9	1657	9	ADB30061	Adb30061 cDNA enco	228	13	92.9	1657	10	ADC58584	Adc58584 Novel hum
156	13	92.9	1657	9	ADA80589	Ad80589 Human PRO	229	13	92.9	1657	10	ADC14498	Adc14498 Novel hum
157	13	92.9	1657	9	ADA75831	Ad75831 Human PRO	230	13	92.9	1657	10	ADD08030	Add08030 Novel hum
158	13	92.9	1657	9	ADA38639	Ad38639 Human cDN	231	13	92.9	1657	10	ADD03258	Add03258 Novel hum
159	13	92.9	1657	9	ADA47056	Ad47056 Human PRO	232	13	92.9	1657	10	ADC90250	Adc90250 Novel hum
160	13	92.9	1657	9	ADB25352	Adb25352 Human PRO	233	13	92.9	1657	10	ADC81855	Adc81855 Human PRO
161	13	92.9	1657	9	ADA93528	Ad93528 Human PRO	234	13	92.9	1657	10	ADC69669	Adc69669 cDNA enco
162	13	92.9	1657	9	ADB26878	Adb26878 cDNA enco	235	13	92.9	1657	10	ADC48558	Adc48558 Human PRO
163	13	92.9	1657	9	ADB31165	Adb31165 cDNA enco	236	13	92.9	1657	10	ADD10087	Add10087 Human PRO
164	13	92.9	1657	9	ADA92760	Ad92760 Human cDN	237	13	92.9	1657	10	ADD07497	Add07497 Novel hum
165	13	92.9	1657	9	ADA61093	Ad61093 Homo sapi	238	13	92.9	1657	10	ADA64662	Ada64662 Novel hum
166	13	92.9	1657	9	ADB24240	Adb24240 Human PRO	239	13	92.9	1657	10	ADC82388	Adc82388 Human PRO

240	13	92.9	1657	10	ADC80618	Adc80618	Novel hum	313	13	92.9	1657	11	ADI63582	Adi63582	Novel hum
241	13	92.9	1657	10	ADD11125	Adc41125	Human PRO	314	13	92.9	1657	12	ADD76618	Adc81170	Novel hum
242	13	92.9	1657	10	ADC48006	Adc48006	Human PRO	315	13	92.9	1657	12	ADD76618	Adc81170	Novel hum
243	13	92.9	1657	10	ADD08568	Adc80666	Novel hum	316	13	92.9	1657	12	ADD87982	Adc86386	Human PRO
244	13	92.9	1657	10	ADC80666	Adc80666	Novel hum	317	13	92.9	1657	12	ADD87982	Adc86386	Human PRO
245	13	92.9	1657	10	ADD06817	Adc06817	Novel hum	318	13	92.9	1657	12	ADe75834	Adc75834	Human PRO
246	13	92.9	1657	10	ADD09535	Adc09535	Human PRO	319	13	92.9	1657	12	ADe23410	Adc23410	CDNA enco
247	13	92.9	1657	10	ADC83064	Adc83064	Human PRO	320	13	92.9	1657	12	ADe23962	Adc23962	CDNA enco
248	13	92.9	1657	10	ADD41248	Adc41248	Novel hum	321	13	92.9	1657	12	ADDe24605	Adc24605	CDNA enco
249	13	92.9	1657	10	ADD52387	Adc52387	CDNA enco	322	13	92.9	1657	12	ADDe24605	Adc24605	CDNA enco
250	13	92.9	1657	10	ADD53127	Adc53127	CDNA enco	323	13	92.9	1657	12	ADDe89296	Adc89296	Human PRO
251	13	92.9	1657	10	ADD53679	Adc53679	Novel hum	324	13	92.9	1657	12	ADDe18435	Adc18435	Human PRO
252	13	92.9	1657	10	ADD55171	Adc55171	Human PRO	325	13	92.9	1657	12	ADDe88744	Adc88744	Human PRO
253	13	92.9	1657	10	ADD56129	Adc56129	Human PRO	326	13	92.9	1657	12	ADDe94764	Adc94764	CDNA enco
254	13	92.9	1657	10	ADD51835	Adc51835	CDNA enco	327	13	92.9	1657	12	ADDe91175	Adc91175	Human PRO
255	13	92.9	1657	10	ADD02634	Adc02634	Human PRO	328	13	92.9	1657	12	ADDF35324	Adf35324	CDNA enco
256	13	92.9	1657	10	ADD02068	Adc02068	Human PRO	329	13	92.9	1657	12	ADDe95316	Adc95316	CDNA enco
257	13	92.9	1657	10	ADD54250	Adc54250	Novel hum	330	13	92.9	1657	12	ADDe93426	Adc93426	Human PRO
258	13	92.9	1657	10	ADD54567	Adc54567	Human PRO	331	13	92.9	1657	12	ADDF35007	Adf35007	CDNA enco
259	13	92.9	1657	10	ADD92567	Adc92567	Human PRO	332	13	92.9	1657	12	ADDe92322	Adc92322	Novel hum
260	13	92.9	1657	10	ADD91463	Adc91463	Human PRO	333	13	92.9	1657	12	ADDe90623	Adc90623	Human PRO
261	13	92.9	1657	10	ADDe04077	Adc04077	Human PRO	334	13	92.9	1657	12	ADDe91770	Adc91770	Novel hum
262	13	92.9	1657	10	ADDe26721	Adc26721	Novel hum	335	13	92.9	1657	12	ADG11574	Adg11574	CDNA enco
263	13	92.9	1657	10	ADDe32374	Adc32374	Novel hum	336	13	92.9	1657	12	ADG02349	Adg02349	Human PRO
264	13	92.9	1657	10	ADDe22306	Adc22306	CDNA enco	337	13	92.9	1657	12	ADG22135	Adg22135	Novel hum
265	13	92.9	1657	10	ADD79530	Adc79530	CDNA enco	338	13	92.9	1657	12	ADG20205	Adg20205	CDNA enco
266	13	92.9	1657	10	ADDe42066	Adc42066	Human PRO	339	13	92.9	1657	12	ADDF98111	Adf98111	Human PRO
267	13	92.9	1657	10	ADDe17883	Adc17883	Human PRO	340	13	92.9	1657	12	ADG24328	Adg24328	Novel hum
268	13	92.9	1657	10	ADDe92015	Adc92015	Human PRO	341	13	92.9	1657	12	ADG98682	Adg98682	Human PRO
269	13	92.9	1657	10	ADDe33478	Adc33478	Novel hum	342	13	92.9	1657	12	ADG03513	Adg03513	Human PRO
270	13	92.9	1657	10	ADDe34030	Adc34030	Novel hum	343	13	92.9	1657	12	ADDF99234	Adf99234	Human PRO
271	13	92.9	1657	10	ADDe80082	Adc80082	CDNA enco	344	13	92.9	1657	12	ADG16819	Adg16819	CDNA enco
272	13	92.9	1657	10	ADD93119	Adc93119	Human PRO	345	13	92.9	1657	12	ADG05278	Adg05278	Human PRO
273	13	92.9	1657	10	ADDe19539	Adc19539	Human PRO	346	13	92.9	1657	12	ADG19545	Adg19545	CDNA enco
274	13	92.9	1657	10	ADDe18987	Adc18987	Human PRO	347	13	92.9	1657	12	ADG13382	Adg13382	CDNA enco
275	13	92.9	1657	10	ADDe43183	Adc43183	Human PRO	348	13	92.9	1657	12	ADG08439	Adg08439	Novel hum
276	13	92.9	1657	10	ADD95972	Adc95972	Human PRO	349	13	92.9	1657	12	ADG15609	Adg15609	CDNA enco
277	13	92.9	1657	10	ADDe22858	Adc22858	CDNA enco	350	13	92.9	1657	12	ADDF97007	Adf97007	Human PRO
278	13	92.9	1657	10	ADD78976	Adc78976	CDNA enco	351	13	92.9	1657	12	ADG06192	Adg06192	Human PRO
279	13	92.9	1657	10	ADDe26188	Adc26188	Novel hum	352	13	92.9	1657	12	ADG23776	Adg23776	Novel hum
280	13	92.9	1657	10	ADDe32926	Adc32926	Novel hum	353	13	92.9	1657	12	ADG04065	Adg04065	Human PRO
281	13	92.9	1657	10	ADDe42618	Adc42618	Human PRO	354	13	92.9	1657	12	ADG24966	Adg24966	Novel hum
282	13	92.9	1657	10	ADD80634	Adc80634	CDNA enco	355	13	92.9	1657	12	ADG07263	Adg07263	Novel hum
283	13	92.9	1657	10	ADDe89662	Adc89662	Human PRO	356	13	92.9	1657	12	ADG07815	Adg07815	Novel hum
284	13	92.9	1657	10	ADDe40946	Adc40946	Human PRO	357	13	92.9	1657	12	ADG55310	Adg55310	Novel hum
285	13	92.9	1657	10	ADDe04745	Adc04745	Human PRO	358	13	92.9	1657	12	ADG60974	Adg60974	Novel hum
286	13	92.9	1657	10	ADDe92874	Adc92874	Human PRO	359	13	92.9	1657	12	ADG62078	Adg62078	Novel hum
287	13	92.9	1657	10	ADDe67125	Adc67125	Human PRO	360	13	92.9	1657	12	ADG82279	Adg82279	Human PRO
288	13	92.9	1657	10	ADG21583	Adg21583	Novel hum	361	13	92.9	1657	12	ADG57518	Adg57518	Novel hum
289	13	92.9	1657	10	ADG23224	Adg23224	Novel hum	362	13	92.9	1657	12	ADG56966	Adg56966	Novel hum
290	13	92.9	1657	10	ADDF97559	Adf97559	Human PRO	363	13	92.9	1657	12	ADG55862	Adg55862	Novel hum
291	13	92.9	1657	10	ADG80623	Adg80623	Human PRO	364	13	92.9	1657	12	ADG58622	Adg58622	Novel hum
292	13	92.9	1657	10	ADG80071	Adg80071	Human PRO	365	13	92.9	1657	12	ADG70988	Adg70988	Novel hum
293	13	92.9	1657	10	ADH55363	Adh55363	Novel hum	366	13	92.9	1657	12	ADG58070	Adg58070	Novel hum
294	13	92.9	1657	10	ADH55915	Adh55915	Novel hum	367	13	92.9	1657	12	ADG53654	Adg53654	Novel hum
295	13	92.9	1657	10	ADH35379	Adh35379	Human PRO	368	13	92.9	1657	12	ADG71540	Adg71540	Novel hum
296	13	92.9	1657	10	ADH64134	Adh64134	Novel hum	369	13	92.9	1657	12	ADH81727	Adh81727	Human PRO
297	13	92.9	1657	10	ADH65083	Adh65083	Novel hum	370	13	92.9	1657	12	ADH19444	Adh19444	Human CDN
298	13	92.9	1657	10	ADH81996	Adh81996	Novel hum	371	13	92.9	1657	12	ADH30689	Adh30689	Human PRO
299	13	92.9	1657	10	ADH99871	Adh99871	Novel hum	372	13	92.9	1657	12	ADH12056	Adh12056	Novel hum
300	13	92.9	1657	10	ADH81444	Adh81444	Novel hum	373	13	92.9	1657	12	ADG52478	Adg52478	Novel hum
301	13	92.9	1657	10	ADH77842	Adh77842	Human PRO	374	13	92.9	1657	12	ADG54206	Adg54206	Novel hum
302	13	92.9	1657	10	ADH80254	Adh80254	Novel hum	375	13	92.9	1657	12	ADG81175	Adg81175	Human PRO
303	13	92.9	1657	10	ACA69160	Adc69160	Human CDN	376	13	92.9	1657	12	ADG56414	Adg56414	Novel hum
304	13	92.9	1657	10	ADDe24069	Adc24069	Novel hum	377	13	92.9	1657	12	ADH12680	Adh12680	Novel hum
305	13	92.9	1657	10	ADH90231	Adh90231	Human sec	378	13	92.9	1657	12	ADH20937	Adh20937	Human CDN
306	13	92.9	1657	10	ADH64077	Adh64077	CDNA enco	379	13	92.9	1657	12	ADG61526	Adg61526	Novel hum
307	13	92.9	1657	10	ADH67210	Adh67210	CDNA enco	380	13	92.9	1657	12	ADH19977	Adh19977	Human CDN
308	13	92.9	1657	11	ADH82613	Adh82613	Novel hum	381	13	92.9	1657	12	ADH28613	Adh28613	Human PRO
309	13	92.9	1657	11	ADH16012	Adh16012	Novel hum	382	13	92.9	1657	12	ADG54758	Adg54758	Novel hum
310	13	92.9	1657	11	ADH16641	Adh16641	Novel hum	383	13	92.9	1657	12	ADG59798	Adg59798	Novel hum
311	13	92.9	1657	11	ADH15460	Adh15460	Novel hum	384	13	92.9	1657	12	ADH181222	Adh181222	CDNA enco
312	13	92.9	1657	11	ADH14908	Adh14908	Novel hum	385	13	92.9	1657	12	ADG09965	Adg09965	Novel hum

386	13	92.9	1657	12	AD115436	Adi15436 Novel hum	459	13	92.9	9990	5	AAH26136	Aah26136 Mouse PrP
387	13	92.9	1657	12	ADG09313	Adg09313 Novel hum	460	13	92.9	12043	6	AAS95418	Aas95418 Human int
388	13	92.9	1657	12	ADI14768	Adi14768 Novel hum	461	13	92.9	12043	6	AAS95361	Aas95361 Human int
389	13	92.9	1657	12	ADI18363	Adi18363 Novel hum	462	13	92.9	17736	4	ABL28554	AbL28554 Drosophil
390	13	92.9	1657	12	ADJ63644	Adj63644 Novel hum	463	13	92.9	18692	4	ABK42795	Abk42795 Genomic s
391	13	92.9	1657	12	ADJ77539	Adj77539 Human PRO	464	13	92.9	18692	9	ADB60951	AdB60951 Connectiv
392	13	92.9	1657	12	ADJ65661	Adj65661 cDNA enco	465	13	92.9	19062	3	AAF21280	Aaf21280 Human low
393	13	92.9	1657	12	ADM27797	Adm27797 cDNA enco	466	13	92.9	19062	10	ABZ96974	Abz96974 Human nuc
394	13	92.9	1657	12	ADM42521	Adm42521 cDNA enco	467	13	92.9	19062	11	ABD20823	Abd20823 Human pul
395	13	92.9	1657	12	ADM28383	Adm28383 cDNA enco	468	13	92.9	20067	4	AAK66735	Aak66735 Human imm
396	13	92.9	1657	12	ADI95865	Adi95865 cDNA enco	469	13	92.9	20068	4	AAK66733	Aak66733 Human imm
397	13	92.9	1657	13	ADI96417	Adi96417 Novel hum	470	13	92.9	20323	4	AAK66731	Aak66731 Human imm
398	13	92.9	1668	9	ADA48431	Ada48431 Rice gene	471	13	92.9	23142	3	AAA35158	Aaa35158 Human ade
399	13	92.9	1677	13	ACN39308	Acn39308 Tumour-as	472	13	92.9	26147	11	ACN44862	Acn44862 Human gen
400	13	92.9	1725	6	ABO72565	Abg72565 Human MDD	473	13	92.9	27315	4	ABL08822	AbL08822 Drosophil
401	13	92.9	1727	6	AAS62254	Aas62254 cDNA sequ	474	13	92.9	27413	11	ACN44832	Acn44832 Mouse gen
402	13	92.9	1767	5	ABA19403	Abal19403 Human ner	475	13	92.9	30417	3	AAA35160	Aaa35160 Human ade
403	13	92.9	1768	5	ABA19404	Abal19404 Human ner	476	13	92.9	30417	3	AAF21282	Aaf21282 Human low
404	13	92.9	1782	5	AAS78516	Aas78516 DNA enco	477	13	92.9	30417	10	ABZ96976	Abz96976 Human nuc
405	13	92.9	1910	2	AAV59790	Aav59790 Human sec	478	13	92.9	30417	11	ABD20825	Abd20825 Human pul
406	13	92.9	1910	6	ABS73784	Abas73784 Human CDN	479	13	92.9	32328	4	ABL09994	AbL09994 Drosophil
407	13	92.9	1910	9	ACD82327	AcD82327 cDNA sequ	480	13	92.9	33454	11	ACN44588	Acn44588 Mouse gen
408	13	92.9	1910	10	ADI23012	Adi23012 cDNA enco	481	13	92.9	40839	4	ABL20436	AbL20436 Drosophil
409	13	92.9	1910	12	ADH74014	Adh74014 Human sec	482	13	92.9	93483	9	ADA03083	Ada03083 Mouse mCG
410	13	92.9	1943	10	ADC23966	Adc23966 Human nov	483	13	92.9	93483	9	ADA66367	Ada66367 Mouse mCG
411	13	92.9	1945	2	AAS59664	Aav59664 Human sec	484	13	92.9	93483	10	ADB72821	AdB72821 Mouse mCG
412	13	92.9	1945	6	ABS73651	Abas73651 Human CDN	485	13	92.9	93483	11	ADL27161	AdL27161 Mouse gen
413	13	92.9	1945	9	ACD82794	AcD82794 cDNA sequ	486	13	92.9	96256	13	ADR53001	Adr53001 Drug ther
414	13	92.9	1945	10	ADI22879	Adi22879 cDNA enco	487	13	92.9	98642	11	ACN44584	Acn44584 Mouse gen
415	13	92.9	1945	12	ADH73881	Adh73881 Human sec	488	13	92.9	100267	6	ABT11032	AbT11032 Human bre
416	13	92.9	2021	10	ADB62295	AdB62295 Human CDN	489	13	92.9	101241	11	ACN44740	Acn44740 Mouse gen
417	13	92.9	2048	5	AAS90533	Aas90533 DNA enco	490	13	92.9	106416	4	ABL18718	AbL18718 Drosophil
418	13	92.9	2058	11	ABD00035	Abd00035 Klebsiell	491	13	92.9	110000	4	AAI95682_41	Continuation (42 o
419	13	92.9	2121	4	ABL22212	AbL22212 Drosophil	492	13	92.9	110000	4	AAI95683_41	Continuation (42 o
420	13	92.9	2214	5	AAS66206	Aas66206 DNA enco	493	13	92.9	115284	11	ACN44296	Acn44296 Mouse gen
421	13	92.9	2242	10	ADD48071	AdD48071 Rat gene	494	13	92.9	126192	11	ACN44820	Acn44820 Mouse gen
422	13	92.9	2263	5	AAS92449	Aas92449 DNA enco	495	13	92.9	130244	13	ABD32872	Abd32872 Human can
423	13	92.9	2468	12	ADP72650	Adp72650 Renal tox	496	13	92.9	131576	11	ACN44890	Acn44890 Human can
424	13	92.9	2474	10	ABT41975	Abt41975 Toxicity	497	13	92.9	197775	11	ACN44416	Acn44416 Mouse gen
425	13	92.9	2475	9	ACF25356	ACF25356 Rat MAG 9	498	13	92.9	264965	12	ADN16203	Adn16203 Human eul
426	13	92.9	2475	10	ADB58278	AdB58278 Toxicity-	499	13	92.9	268685	6	ABS56563	AbS56563 Human SUL
427	13	92.9	2475	12	ADR13958	Adr13958 Rat myeli	500	12.6	90.0	574	4	AAL01368	Aal01368 Human rep
428	13	92.9	2475	13	ADT89534	Adt89534 Rattus no							
429	13	92.9	2540	10	ADB62777	AdB62777 Human CDN							
430	13	92.9	2595	5	AAS12711	Aas12711 Rat bombe							
431	13	92.9	2902	4	ABL17892	AbL17892 Drosophil							
432	13	92.9	3000	8	ABZ36093	Abz36093 Human sec							
433	13	92.9	3259	13	ADB58069	AdB58069 Toxicity-							
434	13	92.9	3259	13	ADT89543	Adt89543 Rattus no							
435	13	92.9	3260	3	AAA09356	Aaa09356 p75-NTR (
436	13	92.9	3386	10	AAD60676	Aad60676 Wheat sta							
437	13	92.9	3530	6	ABL69227	AbL69227 Prostate							
438	13	92.9	3530	6	ABL69756	AbL69756 Prostate							
439	13	92.9	3530	6	ABK84630	AbK84630 Human CDN							
440	13	92.9	3530	6	ABN97276	Abn97276 Gene #377							
441	13	92.9	3645	5	ABV24673	Abv24673 Human pro							
442	13	92.9	3645	5	ABV28517	Abv28517 Human pro							
443	13	92.9	3645	5	ABV22892	Abv22892 Human pro							
444	13	92.9	3645	5	ABV29547	Abv29547 Human pro							
445	13	92.9	3684	10	ADES3897	AdeS3897 Human pro							
446	13	92.9	3914	12	ADQ08599	Adq08599 Ciona int							
447	13	92.9	4567	12	ADO44182	Ado44182 DNA enco							
448	13	92.9	5968	4	AAS27841	Aas27841 DNA enco							
449	13	92.9	5968	4	AAS35091	Aas35091 DNA #41 e							
450	13	92.9	5968	4	AAK80295	Aak80295 Human imm							
451	13	92.9	5968	10	ADB94644	AdB94644 Novel hum							
452	13	92.9	5968	10	ADC46533	Adc46533 Human neo							
453	13	92.9	7231	5	ABA15980	Abal15980 Human ner							
454	13	92.9	7233	5	ABA15981	Abal15981 Human ner							
455	13	92.9	7234	5	ABA15979	Abal15979 Human ner							
456	13	92.9	7505	13	ACN40825	Acn40825 Tumour-as							
457	13	92.9	9190	6	AAL39555	Aal39555 Genomic D							
458	13	92.9	9990	5	AAH26123	Aah26123 Mouse pri							

ALIGNMENTS

RESULT 1

AAAX40730
 ID AAAX40730 standard; DNA; 14 BP.
 XX
 AC AAAX40730;
 XX
 DT 16-JUN-1999 (first entry)
 XX
 GC-box e-MET nucleotide sequence.
 DE
 KW DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
 KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
 KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
 XX ss.
 OS Synthetic.
 XX
 FN WO9912027-A1.
 XX
 PD 11-MAR-1999.
 XX
 PF 12-JUN-1998; 98WO-US012351.
 XX
 PR 29-AUG-1997; 97US-0057411P.
 XX
 PA (REGC) UNIV CALIFORNIA.

Reich NO, Flynn J;
WPI; 1999-205256/17.
DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
used to inhibit methylation of DNA, and proliferation of cancer cells.
Claim 10; Page 63; 114pp; English.
The specification describes a synthetic oligonucleotide comprising a C-5
methylcytosine which recognizes and binds an allosteric site on DNA
cytosine methyltransferase (DNMTase), thereby modulating DNMTase
activity associated with the allosteric site. The synthetic inhibitor can
be used to inhibit methylation of DNA. It can also be used to inhibit
proliferation of cancer cells. The inhibitor can also be used to treat a
disorder of development, which is linked to a genetic locus regulated by
methylation, such as Huntington's disease, Down's syndrome, and disorders
associated with a Hox gene. The present sequence appears in the
specification

Sequence 14 BP; 0 A; 8 C; 2 G; 4 T; 0 U; 0 Other;
Query Match 100.0%; Score 14; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 TTGCCCGCCGCTT 14
Db 1 TTGCCCGCCGCTT 14

RESULT 2
AAX40729
ID AAX40729 standard; DNA; 22 BP.
XX
AC AAX40729;
XX
DT 16-JUN-1999 (first entry)
XX
DE GC-box d-MET nucleotide sequence.
XX
KW DNA cytosine methyltransferase; DNMTase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW ss.
XX
OS Synthetic.
XX
PN WO9912027-A1.
XX
PD 11-MAR-1999.
XX
PF 12-JUN-1998; 98WO-US012351.
XX
PR 29-AUG-1997; 97US-0057411P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Reich NO, Flynn J;
XX
PS WPI; 1999-205256/17.
XX
PT DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
PT used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
Claim 10; Page 63; 114pp; English.
XX
The specification describes a synthetic oligonucleotide comprising a C-5
XX methylcytosine which recognizes and binds an allosteric site on DNA
XX cytosine methyltransferase (DNMTase), thereby modulating DNMTase
XX activity associated with the allosteric site. The synthetic inhibitor can
XX be used to inhibit methylation of DNA. It can also be used to inhibit
XX proliferation of cancer cells. The inhibitor can also be used to inhibit
XX disorder of development, which is linked to a genetic locus regulated by
XX methylation, such as Huntington's disease, Down's syndrome, and disorders
XX associated with a Hox gene. The present sequence appears in the
XX specification

```

RESULT 4
AAx40724/c
ID AAX40724 standard; DNA; 30 BP.
XX
AC AAX40724;
XX
DT 16-JUN-1999 (first entry)
XX
DE GC-box a nucleotide sequence of the specification.
XX
KW DNA cytosine methyltransferase; DCMtase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW ss.
XX
OS Synthetic.
XX
PN WO9912027-A1.
XX
PD 11-MAR-1999.
XX
PF 12-JUN-1998; 98WO-US012351.
XX
PR 29-AUG-1997; 97US-0057411P.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Reich NO, Flynn J;
XX
WPI; 1999-205256/17.
XX
DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
Claim 10; Page 63; 114pp; English.
XX
The specification describes a synthetic oligonucleotide comprising a C-5
methylcytosine which recognizes and binds an allosteric site on DNA
cytosine methyltransferase (DCMTase) , thereby modulating DCMtase
activity associated with the allosteric site. The synthetic inhibitor can
be used to inhibit methylation of DNA. It can also be used to inhibit
proliferation of cancer cells. The inhibitor can also be used to treat a
disorder of development, which is linked to a genetic locus regulated by
methylation, such as Huntington's disease, Down's syndrome, and disorders
associated with a Hox gene. The present sequence appears in the
specification
XX
Sequence 30 BP; 8 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 14; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
Db |||||
22 TTGCCCCGCCCTT 9
RESULT 5
AAx40728
ID AAX40728 standard; DNA; 50 BP.
XX
AC AAX40728;
XX
DT 16-JUN-1999 (first entry)
XX
DE GC-box C-MET nucleotide sequence.
XX
KW DNA cytosine methyltransferase; DCMtase; C-5 methylcytosine;
KW allosteric site; synthetic inhibitor; cancer cell; proliferation;
KW development disorder; Huntington's disease; Down's syndrome; Hox gene;
KW

```

```

KW ss.
XX
OS Synthetic.
XX
PN WO9912027-A1.
XX
PD 11-MAR-1999.
XX
PF 12-JUN-1998; 98WO-US012351.
XX
PR 29-AUG-1997; 97US-0057411P.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Reich NO, Flynn J;
XX
WPI; 1999-205256/17.
XX
DNA cytosine methyltransferase modulator containing 5-C methylcytosine -
used to inhibit methylation of DNA, and proliferation of cancer cells.
XX
Claim 10; Page 63; 114pp; English.
XX
The specification describes a synthetic oligonucleotide comprising a C-5
methylcytosine which recognizes and binds an allosteric site on DNA
cytosine methyltransferase (DCMTase) , thereby modulating DCMtase
activity associated with the allosteric site. The synthetic inhibitor can
be used to inhibit methylation of DNA. It can also be used to inhibit
proliferation of cancer cells. The inhibitor can also be used to treat a
disorder of development, which is linked to a genetic locus regulated by
methylation, such as Huntington's disease, Down's syndrome, and disorders
associated with a Hox gene. The present sequence appears in the
specification
XX
Sequence 50 BP; 8 A; 27 C; 5 G; 10 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 14; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
Db |||||
19 TTGCCCCGCCCTT 32
RESULT 6
ACD98140/c
ID ACD98140 standard; cDNA; 255 BP.
XX
AC ACD98140;
XX
DT 23-SEP-2003 (first entry)
XX
DE Human colon cancer cell expressed cDNA #6552.
XX
KW Open reading frame detection; genome sequencing; colon cancer;
KW breast cancer; population genome analysis; genetic shift; cancer;
KW antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW agriculture; food crop genome; resistance gene; retrovirus;
KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN US2002155438-A1.
XX
PD 24-OCT-2002.
XX
PF 27-SEP-1999; 99US-00406117.
XX
PR 20-NOV-1998; 98US-00196716.
XX
PA (SIMP/) SIMPSON A J G.

```

PA (NETO/) NETO E D.
 XX (BREN/) BRENTANI R R.
 PI Simpson AJG, Neto ED, Brentani RR;
 XX WPI; 2003-182626/18.
 DR
 XX
 PT Determining open reading frames of genome of an organism e.g. a human
 PT suffering from cancer involves use of single oligonucleotide primer at
 PT low stringency for preparing single-stranded cDNA from mRNA of
 PT individual.
 XX
 XX Example 9; Page 928; 959pp; English.
 PS
 XX
 CC The invention describes a method of determining open reading frames in
 CC the genome of organism, comprising contacting mRNA from cell of organism
 CC with a single oligonucleotide primer (I) at low stringency, preparing
 CC single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
 CC cDNA, sequencing the product, and repeating the contacting, preparing
 CC and amplifying steps with different primers and sequencing resulting
 CC nucleic acids. The method is useful for: determining that a known
 CC nucleotide sequence from a genome of an organism corresponds to a
 CC nucleotide sequence of an open reading frame; for preparing a contig,
 CC nucleic acid molecule from a genome of an organism; and for sequencing
 CC all or part of a genome of an organism. mRNA is obtained from mammalian
 CC or human cell which is associated with a pathological condition e.g. a
 CC colon cancer or breast cancer cell. The method is useful for analyses of
 CC populations of subjects and can be used to carry out genetic analyses of
 CC large or small populations. further, it can be used to study living
 CC systems to determine if, e.g. there have been genetic shifts which render
 CC an individual or population more or less likely to be afflicted with
 CC diseases such as cancer, to determine antibiotic resistance or non-
 CC tolerance, and so forth. The method can also be used in the study of
 CC congenital diseases, and the risk of affliction to a foetus, as well as
 CC the study of whether the conditions are likely to be passed to offspring
 CC through ova or sperm. The analyses for pathological conditions can be
 CC carried out in all animals, plants, birds, fish, etc. Using this method,
 CC in the area of agriculture, for example the genomes of food crops can be
 CC studied to determine if resistance genes are present, defects in plant
 CC genomes can also be studied in this way. Similarly, the method permits
 CC determination of the pathogens which integrate into the genome, such as
 CC retroviruses and other integrating viruses such as influenza virus, have
 CC undergone shifts or mutations, which may require different approaches to
 CC therapy. This method is also applied to eukaryotic pathogens, such as
 CC trypanosomes, different types of Plasmodium, etc. The method essentially
 CC eliminates sequencing of non-coding portions. This sequence represents a
 CC polynucleotide isolated from human colon cancer cell cDNA library
 XX
 SQ Sequence 255 BP; 59 A; 56 C; 80 G; 60 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 10; Length 255;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 203 TTGCCCCGCCCTT 190
 |||||
 RESULT 7
 ID ABX35297 standard; cDNA; 292 BP.
 XX ABX35297;
 AC
 XX 20-FEB-2003 (first entry)
 DT
 XX Bovine EST associated with lactation/muscle/fat deposition #462.
 DE
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX

OS Bos Taurus.
 XX US2002137139-A1.
 PN
 XX
 PD 26-SEP-2002.
 XX
 PF 24-SEP-2001; 2001US-00960352.
 XX
 XX 12-JAN-1999; 99US-0115707P.
 PR 11-JAN-2000; 2000US-00480902.
 XX
 XX (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 WPI; 2003-110599/10.
 DR
 XX
 XX New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 PT cattle breeding, or for genetically improving cattle.
 PT
 XX Claim 2; SEQ ID NO 462; 245pp; English.
 PS
 XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non- translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?docID=20020137139
 XX
 SQ Sequence 292 BP; 58 A; 110 C; 52 G; 72 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 8; Length 292;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 51 TTGCCCCGCCCTT 64
 |||||
 RESULT 8
 ID AAS39325/c standard; cDNA; 326 BP.
 XX AAS39325;
 AC
 XX 17-DEC-2001 (first entry)
 DT
 XX Novel human diagnostic and therapeutic gene #2383.
 DE

```

XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
KW Homo sapiens.
OS WO200166753-A2.
XX
XX 13-SEP-2001.
XX
XX 09-MAR-2001; 2001WO-US007787.
XX
XX 09-MAR-2000; 2000US-0188609P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
XX WPI; 2001-530177/58.
XX
XX New polynucleotides and polypeptides, useful for diagnosis and treatment
PT of breast, lung and colon cancer.
XX
XX Claim 1; Page 1190; 1193pp; English.
XX
XX The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention
XX
XX Sequence 326 BP; 87 A; 67 C; 90 G; 82 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 14; DB 4; Length 326;
XX Best Local Similarity 100.0%; Pred. No. 1.5e+03;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTGCCCCGCCCTT 14
XX |||||||||
XX Db 128 TTGCCCCGCCCTT 115
XX
XX RESULT 9
XX AAS39329/c
XX ID AAS39329 standard; cDNA; 340 BP.
XX
XX AC AAS39329;
XX
XX DT 17-DEC-2001 (first entry)
XX
XX DE Novel human diagnostic and therapeutic gene #2387.
XX
XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
KW Homo sapiens.
XX
XX WO200166753-A2.
XX
XX 13-SEP-2001.
XX
XX 09-MAR-2001; 2001WO-US007787.
XX
XX 09-MAR-2000; 2000US-0188609P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.

```

```

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
XX WPI; 2001-530177/58.
XX
XX New polynucleotides and polypeptides, useful for diagnosis and treatment
PT of breast, lung and colon cancer.
XX
XX Claim 1; Page 1191; 1193pp; English.
XX
XX The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention
XX
XX Sequence 340 BP; 89 A; 70 C; 95 G; 86 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 14; DB 4; Length 340;
XX Best Local Similarity 100.0%; Pred. No. 1.5e+03;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTGCCCCGCCCTT 14
XX |||||||||
XX Db 128 TTGCCCCGCCCTT 115
XX
XX RESULT 10
XX AAF65653/c
XX ID AAF65653 standard; cDNA; 411 BP.
XX
XX AC AAF65653;
XX
XX DT 09-APR-2001 (first entry)
XX
XX DE Novel human polynucleotide, SEQ ID NO: 1409.
XX
XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200102568-A2.
XX
XX PD 11-JAN-2001.
XX
XX PF 30-JUN-2000; 2000WO-US018374.
XX
XX PR 02-JUL-1999; 99US-0142310P.
XX PR 02-JUL-1999; 99US-0142311P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI Drmanac R, Crkvenjakov R, Drmanac S, Dickson M, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
XX WPI; 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences.
XX
XX Claim 9; Page 746; 1046pp; English.

```

```
XX CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia
XX SQ Sequence 411 BP; 90 A; 111 C; 113 G; 97 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 5; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
DB 148 TTGCCCCGCCCTT 135

RESULT 11
AAS34163
ID AAS34163 standard; cDNA; 587 BP.
AC AAS34163;
XX 17-DEC-2001 (first entry)
DT
XX Human cDNA encoding a novel foetal antigen, SEQ ID No 687.
DE
XX Human; foetal tissue antigen; as; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
XX WO200155312-A2.
PN
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001321.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
```



```
QY      1 TTGCCCCGCCCTT 14
DB      120 TTGCCCCGCCCTT 133

RESULT 13
ABL04361
ID ABL04361 standard; cDNA; 1062 BP.
XX
AC ABL04361;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7565.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB60258.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 7565; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1062 BP; 265 A; 267 C; 283 G; 247 T; 0 U; 0 Other;

Query Match      100.0%; Score 14; DB 4; Length 1062;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
DB      602 TTGCCCCGCCCTT 615

RESULT 14
AAS65178/c
ID AAS65178 standard; cDNA; 1359 BP.
XX
AC AAS65178;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #982.
```

```
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABB00991.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 982; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1359 BP; 322 A; 325 C; 402 G; 310 T; 0 U; 0 Other;

Query Match      100.0%; Score 14; DB 5; Length 1359;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
DB      714 TTGCCCCGCCCTT 701

RESULT 15
AAS53007/c
ID AAS53007 standard; cDNA; 1563 BP.
XX
AC AAS53007;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #150.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
```

KW treatment; ds.
 XX Homo sapiens.
 OS DE19820190-A1.
 XX PD 04-NOV-1999.
 XX PF 28-APR-1998; 98DE-01020190.
 XX PR 28-APR-1998; 98DE-01020190.
 PA (META-) METAGEN CBS GENOMFORSCHUNG MBH.
 XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 DR WPI; 1999-621386/54.
 XX New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins.
 XX PS Claim 2; Page 301; 502pp; German.
 XX This invention describes novel polypeptides and their encoding nucleic
 CC acids derived from human pancreatic tumor tissue which have cytostatic
 CC activity. The sequences are also useful in producing pharmaceutical
 CC compositions for treatment of pancreatic tumors. AAZ52858-253014
 CC represent expressed sequence tag (EST) fragments derived from a human
 CC pancreatic tumor cDNA library and which encode the proteins represented
 CC in AAY73814-Y74252
 XX
 SQ Sequence 1563 BP; 384 A; 375 C; 461 G; 343 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 2; Length 1563;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 |||||
 DB 653 TTGCCCCGCCCTT 640
 RESULT 16
 AA565180/c
 ID AA565180 standard; cDNA; 1597 BP.
 XX AC AA565180;
 XX DT 13-FEB-2002 (first entry)
 XX DE DNA encoding novel human diagnostic protein #984.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 FN 11-OCT-2001.
 XX PD 30-MAR-2001; 2001WO-US008631.
 XX PF 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG00993.
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 984; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1597 BP; 370 A; 389 C; 468 G; 370 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 5; Length 1597;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 |||||
 DB 714 TTGCCCCGCCCTT 701
 RESULT 17
 ACN40708/c
 ID ACN40708 standard; cDNA; 1597 BP.
 XX AC ACN40708;
 XX DT 18-NOV-2004 (first entry)
 XX DE Tumour-associated antigenic target (TAT) cDNA DNA326782, SEQ ID NO:5661.
 XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 KW central nervous system cancer; bladder cancer; pancreatic cancer;
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;
 KW chromosome identification; chromosome mapping; gene mapping;
 KW gene therapy; cytostatic; gene; ss.
 XX Homo sapiens.
 OS WO2004030615-A2.
 XX PN 15-APR-2004.
 XX PD 29-SEP-2003; 2003WO-US028547.
 XX PF 02-OCT-2002; 2002US-0414971P.
 XX PR (GETH) GENENTECH INC.
 XX Wu TD, Zhang Z, Zhou Y;
 XX WPI; 2004-347921/32.
 DR

DR P-PSDB; ABM02197.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
XX Claim 1; SEQ ID NO 5661; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
XX Sequence 1597 BP; 370 A; 388 C; 471 G; 368 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 14; DB 13; Length 1597;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCCGCCCTT 14
DB 714 TTGCCCCCGCCCTT 701
RESULT 18
ADR66182/c
ID ADR66182 standard; DNA; 1597 BP.
XX AC ADR66182;
XX
XX 02-DEC-2004 (first entry)
XX
XX Human prostatic carcinoma derived DNA SEQ ID 36 #2.
XX
XX human; cytostatic; diagnosis; prostatic cancer;
KW differential expression analysis; ds.
XX
XX Homo sapiens.
XX
XX WO2004076614-A2.
XX
XX 10-SEP-2004.
XX
XX 22-FEB-2004; 2004WO-DE000433.
XX
XX 27-FEB-2003; 2003DE-01009985.
PR 14-MAY-2003; 2003DE-01022134.
XX
XX (HINZ/) HINZMANN B.
PA (DAHL/) DAHL E.
PA (ROSE/) ROSENTHAL A.
PA (HERM/) HERMANN K.
PA (PILA/) PILARSKY C.
XX
XX Hinzmann B, Dahl E, Rosenthal A, Hermann K, Pilarsky C, Specht T;
PI Schmitt A, Beckmann G, Bruemmendorf T, Kinnemann H, Roepcke S;

PI Xinzhong L, Staub E;
XX
XX WPI; 2004-653386/63.
XX
XX New nucleic acids, and encoded proteins, from prostatic cancer tissue,
PT useful for diagnosis, treatment and in screening for specific binding
PT agents.
XX
XX Claim 1; Page 439; 1607pp; German.
XX
XX This invention describes novel cytostatic polynucleotide and polypeptide
CC sequences which can be used in a method for diagnosing prostatic cancer
CC or the risk of developing prostatic cancer. Diagnosis is based on
CC determining over transcription or over expression of the sequences in
CC prostatic tissue. Screening for inhibitors of the sequences or detection
CC substances involves a binding assay, any compounds that bind are
CC selected, optionally after deconvolution of mixtures. Detection of a
CC predetermined minimum level of the reporter indicates the presence of
CC tumour cells. Inhibitors can be chosen from antisense oligonucleotides,
CC short-interfering RNA or ribozymes; an organic molecule of molecular
CC weight below 5000, preferably 300, that binds to the polypeptide; an
CC aptamer against the polypeptide; a (monoclonal) antibody (Ab) against the
CC polypeptide, preferably humanised or human; an anti-idiotypic, non-human
CC (monoclonal) antibody directed against Ab or any of the above derivatised
CC with a reporter group, cell toxin, immunostimulatory molecules and/or
CC radioisotope. The polynucleotides are identified in human prostatic
CC cancer by differential expression analysis, using DNA microarrays,
CC between normal and tumorous tissues, with (over)expression being detected
CC by quantitative PCR. Analysis of prostatic cancer samples showed that
CC CD24 was upregulated in many of them. Sections of tissue, isolated from
CC prostatic cancer patients, or subjects at risk, were incubated
CC sequentially with anti-human CD4 murine monoclonal antibodies;
CC biotinylated second antibody; streptavidin-conjugated horseradish
CC peroxidase and then diaminobenzidine as colour former (brown). The
CC samples were counterstained with hemalum (blue). Malignant cells stained
CC strongly but non-malignant cells only weakly. In 15 of 63 samples of
CC adenocarcinoma, membrane and cytoplasmic staining was very strong, and
CC lymph node metastases were also stained. ADR65805-ADR66954 represent the
CC polynucleotide and polypeptide sequences used in the method of the
XX invention.
XX
XX Sequence 1597 BP; 370 A; 388 C; 471 G; 368 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 14; DB 13; Length 1597;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCCGCCCTT 14
DB 714 TTGCCCCCGCCCTT 701
RESULT 19
ADQ84817/c
ID ADQ84817 standard; cDNA; 1619 BP.
XX AC ADQ84817;
XX
XX 07-OCT-2004 (first entry)
XX
XX Human tumour-associated antigenic target (TAT) cDNA sequence #1631.
XX
XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KW cancer; cell proliferative disorder; gene; ss.
XX
XX Homo sapiens.
XX
XX WO2004060270-A2.
XX
XX 22-JUL-2004.
XX
XX 15-OCT-2003; 2003WO-US029126.
XX

PR 18-OCT-2002; 2002US-0418988P.
XX (GETH) GENENTECH INC.
PA (WUTD//) WU T D.
PA (ZHOU//) ZHOU Y.
XX
PI Wu TD, Zhou Y;
XX
XX WPI; 2004-534300/51.
XX
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 1631; 5504pp; English.
XX
CC The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
XX human TAT cDNA sequence from the present invention.
SQ Sequence 1619 BP; 368 A; 396 C; 490 G; 365 T; 0 U; 0 Other;
Query Match 100.0%; Score 14; DB 12; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
|||||
Db 742 TTGCCCCGCCCTT 729
RESULT 20
ADQ85892/c
ID ADQ85892 standard; cDNA; 1619 BP.
XX
AC ADQ85892;
XX
XX 07-OCT-2004 (first entry)
XX

DE Human tumour-associated antigenic target (TAT) cDNA sequence #2764.
XX
KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KW cancer; cell proliferative disorder; gene; ss.
XX
OS Homo sapiens.
XX
XX WO2004060270-A2.
XX
XX 22-JUL-2004.
XX
XX 15-OCT-2003; 2003WO-US029126.
XX
XX 18-OCT-2002; 2002US-0418988P.
XX
XX (GETH) GENENTECH INC.
PA (WUTD//) WU T D.
PA (ZHOU//) ZHOU Y.
XX
PI Wu TD, Zhou Y;
XX
XX WPI; 2004-534300/51.
XX
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 2764; 5504pp; English.
XX
CC The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
XX human TAT cDNA sequence from the present invention.
SQ Sequence 1619 BP; 368 A; 396 C; 490 G; 365 T; 0 U; 0 Other;
Query Match 100.0%; Score 14; DB 12; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db |||||
742 TTGCCCCGCCCTT 729

RESULT 21
AAH14860/c
ID AAH14860 standard; cDNA; 1622 BP.
XX
AC AAH14860;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12700.
XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

PA

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX

XX Claim 9; SEQ ID NO 12700; 2537pp + Sequence Listing; English.

XX

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dT primer and an oligonucleotide which comprises one of the 5602

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

XX present invention

XX

SQ Sequence 1622 BP; 370 A; 396 C; 491 G; 365 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 4; Length 1622;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db |||||
745 TTGCCCCGCCCTT 732

RESULT 22
ADR21626/c
ID ADR21626 standard; cDNA; 1655 BP.
XX
AC ADR21626;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human enzyme ENZM-8 encoding cDNA SEQ ID NO:47.

KW human; enzyme; ENZM-8; immunosuppressive; antiinflammatory;
KW antimicrobial; neuroprotective; cardiovascular; ophthalmological;
KW gynaecological; cytostatic; gene therapy; immune deficiency;
KW autoimmune disorder; inflammatory disorder; infectious disorder;
KW neurological disorder; cardiovascular disorder; eye disorder;
KW metabolic disorder; reproductive disorder; cell proliferative disorder;
KW cancer; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 180..1247

XX /*tag= a

XX /product= "enzyme ENZM-8"

XX PN WO2004072267-A2.

XX

XX 26-AUG-2004.

XX

XX 12-FEB-2004; 2004WO-US004280.

XX

XX 12-FEB-2003; 2003US-0447246P.

XX 21-FEB-2003; 2003US-0449087P.

XX 26-FEB-2003; 2003US-0450622P.

XX 21-MAR-2003; 2003US-0456704P.

XX 15-APR-2003; 2003US-0463194P.

XX 09-MAY-2003; 2003US-0469358P.

XX 02-JUN-2003; 2003US-0475532P.

XX 04-JUN-2003; 2003US-0476278P.

XX 27-JUN-2003; 2003US-0483395P.

XX (INCY-) INCYTE CORP.

XX

XX Kable AE, Yue H, Baughn MR, Tribouley CM, Ring HZ, Tran UK;

XX Emerling BM, Ramkumar J, Hafalia AJA, Swarnakar A, Lee SY;

XX Chawla NK, Gietzen KJ, Marquis JP, Elliott VS, Becha SD, Favero KD;

XX Wang JT, Naidu S, Hawkins PR, Jin P, Chien D;

XX WPI; 2004-625866/60.

XX P-PSDB; ADR21587.

XX

XX New enzymes, useful for diagnosing, treating, or preventing autoimmune,

XX inflammatory, infectious, neurological, cardiovascular, eye, metabolic,

XX reproductive, or cell proliferative disorders including cancer.

XX

XX Claim 5; SEQ ID NO 47; 305pp; English.

XX

XX The present sequence encodes a human enzyme designated ENZM-8. Human ENZM

XX sequences have immunosuppressive, antiinflammatory, antimicrobial,

XX neuroprotective, cardiovascular, ophthalmological, gynaecological and

XX cytoskeletal activities, and can be used in gene therapy. The human ENZM

XX polypeptides, polynucleotides, compositions, and methods of the present

XX invention can be used for diagnosing, treating, or preventing immune

XX deficiencies, or autoimmune, inflammatory, infectious, neurological,

XX cardiovascular, eye, metabolic, reproductive, or cell proliferative

XX disorders including cancer (e.g. breast, lung, colon, or ovarian cancer).

XX

```
SQ Sequence 1655 BP; 426 A; 399 C; 480 G; 350 T; 0 U; 0 Other;
Query Match 100.0%; Score 14; DB 13; Length 1655;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 724 TTGCCCCGCCCTT 711

RESULT 23
AAS79453
ID AAS79453 standard; cDNA; 2064 BP.
XX
AC AAS79453;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15257.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG15266.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 15257; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2064 BP; 440 A; 569 C; 558 G; 496 T; 0 U; 1 Other;
Query Match 100.0%; Score 14; DB 5; Length 2064;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 1309 TTGCCCCGCCCTT 1322

RESULT 24
ABT42316
ID ABT42316 standard; DNA; 2573 BP.
XX
AC ABT42316;
XX
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 2018.
XX
KW Toxic effect; Gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.
XX
OS Rattus norvegicus.
XX
XX WO200295000-A2.
XX
XX 28-NOV-2002.
XX
XX 22-MAY-2002; 2002WO-US016173.
XX
XX 22-MAY-2001; 2001US-0292335P.
PR
XX 13-JUN-2001; 2001US-0297523P.
PR
XX 19-JUN-2001; 2001US-0298925P.
PR
XX 10-JUL-2001; 2001US-0303807P.
PR
XX 10-JUL-2001; 2001US-0303808P.
PR
XX 10-JUL-2001; 2001US-0303810P.
PR
XX 28-AUG-2001; 2001US-0315047P.
PR
XX 27-SEP-2001; 2001US-0324928P.
PR
XX 22-OCT-2001; 2001US-0330462P.
PR
XX 01-NOV-2001; 2001US-0330867P.
PR
XX 21-NOV-2001; 2001US-0331805P.
PR
XX 06-DEC-2001; 2001US-0336144P.
PR
XX 19-DEC-2001; 2001US-0340873P.
PR
XX 21-FEB-2002; 2002US-0357842P.
PR
XX 21-FEB-2002; 2002US-0357843P.
PR
XX 15-MAR-2002; 2002US-0357844P.
PR
XX 08-APR-2002; 2002US-0364134P.
PR
XX 08-APR-2002; 2002US-0370144P.
PR
XX 08-APR-2002; 2002US-0370206P.
PR
XX 17-APR-2002; 2002US-0370247P.
PR
XX 21-APR-2002; 2002US-0372794P.
XX
XX (GENE-) GENE LOGIC INC.
PA
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
XX Example 4; Page; 446pp; English.
PS
XX The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the
CC progression of a toxic effect of a compound, predicting the renal
```

CC toxicity of a compound, or identifying toxicity markers in tissues or
 CC cells exposed to known renal toxin. The genes are useful as toxicity
 CC markers in drug screening and toxicity assays, in monitoring disease or
 CC physiological states, or disease progression. This polynucleotide
 CC represents a rat DNA sequence relating to the toxic effect database
 CC described in the specification. NOTE: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the World Intellectual Property
 CC Organization
 XX
 SQ Sequence 2573 BP; 672 A; 613 C; 653 G; 635 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 10; Length 2573;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 490 TTGCCCCGCCCTT 503

RESULT 25

ABA05737
 ID ABA05737 standard; cDNA; 3170 BP.

XX
 AC ABA05737;

XX
 DT 04-MAR-2002 (first entry)

XX Murine neuronal serine threonine protein kinase 9B5 coding sequence.

DE Mouse; neuronal serine threonine protein kinase; apoplexy; cytostatic;
 KW cerebroprotective; neuroprotective; antiparkinsonian; immunosuppressive;
 KW viricide; anti-HIV; antiarthritic; antiinflammatory; antiulcer; cancer;
 KW cerebral infarction; neurological disease; immunological disease; atopy;
 KW infection; leukaemia; polyarthritis; Crohn's disease; ulcerative colitis;
 KW 9B5; 88.

XX Mus musculus.

XX Key Location/Qualifiers
 FH CDS 1..2175

FT /*tag= a
 FT /product= "9B5"
 FT polyA_signal 3094..3099
 FT /*tag= b

XX WO200188108-A1.

PN 22-NOV-2001.

XX 17-MAY-2001; 2001WO-EP005660.

XX 17-MAY-2000; 2000DE-01024171.

XX (BADI) BASF-LYNX BIOSCIENCE AG.

XX Schneider A, Klausner B, Fischer A, Newrzella D, Goetz B;

PI Rosner M, Eisenhardt G, Kuner R, Trutzel A, Kammandel B;

PI Jomana Naim S, Schwaninger M;

XX WPI; 2002-055696/07.

DR P-PSDB; ABB04431.

XX New polynucleotide, useful for the diagnosis, treatment and prevention of
 PT e.g. apoplexy, tumors and autoimmune disease, comprises a polynucleotide
 PT encoding the neuronal protein kinase.

XX Claim 1; Page 52-53; 75pp; German.

XX The present invention provides the protein and coding sequences of two
 CC neuronal serine threonine protein kinases each from the mouse and human.
 CC These are designated 9B5 and 9B5_b. Detection of their expression levels

CC can be used to diagnose the risk of apoplexy and in prognosis of cerebral
 CC infarction. They can also be used to identify agents useful in the
 CC prevention, treatment and diagnosis of neurological diseases, including
 CC stroke, multiple sclerosis, Parkinson's disease, amyotrophic lateral
 CC sclerosis, tumours, especially carcinoma, immunological disorders,
 CC including autoimmune diseases, atopy, viral (including human immune
 CC deficiency virus) infections, acute or chronic leukocytic or myeloid
 CC leukaemia, primary chronic polyarthritis, Crohn's disease and ulcerative
 CC colitis. The present sequence is the murine 9B5 coding sequence
 XX
 SQ Sequence 3170 BP; 714 A; 988 C; 897 G; 571 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 6; Length 3170;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 2717 TTGCCCCGCCCTT 2730

RESULT 26

ABA05738

ID ABA05738 standard; cDNA; 3250 BP.

XX
 AC ABA05738;

XX
 DT 04-MAR-2002 (first entry)

DE Murine neuronal serine threonine protein kinase 9B5_b coding sequence.

XX Mouse; neuronal serine threonine protein kinase; apoplexy; cytostatic;
 KW cerebroprotective; neuroprotective; antiparkinsonian; immunosuppressive;
 KW viricide; anti-HIV; antiarthritic; antiinflammatory; antiulcer; cancer;
 KW cerebral infarction; neurological disease; immunological disease; atopy;
 KW infection; leukaemia; polyarthritis; Crohn's disease; ulcerative colitis;
 KW 9B5_b; 88.

XX Mus musculus.

XX Key Location/Qualifiers
 FH CDS 1..1983

FT /*tag= a
 FT /product= "9B5_b"
 FT polyA_signal 3174..3179
 FT /*tag= b

XX WO200188108-A1.

PN 22-NOV-2001.

XX 17-MAY-2001; 2001WO-EP005660.

XX 17-MAY-2000; 2000DE-01024171.

XX (BADI) BASF-LYNX BIOSCIENCE AG.

XX Schneider A, Klausner B, Fischer A, Newrzella D, Goetz B;

PI Rosner M, Eisenhardt G, Kuner R, Trutzel A, Kammandel B;

PI Jomana Naim S, Schwaninger M;

XX WPI; 2002-055696/07.

DR P-PSDB; ABB04432.

XX New polynucleotide, useful for the diagnosis, treatment and prevention of
 PT e.g. apoplexy, tumors and autoimmune disease, comprises a polynucleotide
 PT encoding the neuronal protein kinase.

XX Claim 1; Page 53-54; 75pp; German.

XX The present invention provides the protein anoding sequences of two
 CC neuronal serine threonine protein kinases each from the mouse and human.
 CC These are designated 9B5 and 9B5_b. Detection of their expression levels

CC can be used to diagnose the risk of apoplexy and in prognosis of cerebral
 CC infection. They can also be used to identify agents useful in the
 CC prevention, treatment and diagnosis of neurological diseases, including
 CC stroke, multiple sclerosis, Parkinson's disease, amyotrophic lateral
 CC sclerosis, tumours, especially carcinoma, immunological disorders,
 CC including autoimmune diseases, atopy, viral (including human immune
 CC deficiency virus) infections, acute or chronic leukocytic or myeloid
 CC leukaemia, primary chronic polyarthritis, Crohn's disease and ulcerative
 CC colitis. The present sequence is the murine 985_b coding sequence
 XX
 SQ Sequence 3250 BP; 731 A; 1017 C; 914 G; 588 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 6; Length 3250;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 2797 TTGCCCCGCCCTT 2810

RESULT 27
 ABL11522
 ID ABL11522 standard; cDNA; 3665 BP.
 AC ABL11522;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29048.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 KW Drosophila melanogaster.
 OS

XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
 DR P-PSDB; ABB67419.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX

PS Claim 1; SEQ ID NO 29048; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3665 BP; 926 A; 974 C; 967 G; 798 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 4; Length 3665;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 933 TTGCCCCGCCCTT 946

RESULT 28
 ABL09210/c
 ID ABL09210 standard; cDNA; 4226 BP.

XX ABL09210;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 22112.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB65107.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 22112; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4226 BP; 1192 A; 905 C; 971 G; 1158 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 4; Length 4226;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 2736 TTGCCCCGCCCTT 2723

RESULT 29
 ABL04360
 ID ABL04360 standard; cDNA; 4878 BP.

XX ABL04360;

XX 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7562.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW KM pharmaceutical; gene; ss.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PP 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR P-PSDB; ABB60257.
 XX DR WPI; 2001-656860/75.
 XX DR P-PSDB; ABB60257.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX PS Claim 1; SEQ ID NO 7562; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 4878 BP; 1772 A; 853 C; 805 G; 1448 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 4; Length 4878;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 DB 2009 TTGCCCCGCCCTT 2022
 RESULT 30
 ADO20320
 ID ADO20320 standard; cDNA; 8570 BP.
 XX AC ADO20320;
 XX DT 12-AUG-2004 (first entry)
 XX DE Human PRO polynucleotide #610.
 XX KW Human; PRO; Gene; ss; immune related disorder;
 KW systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis;
 KW juvenile chronic arthritis; systemic sclerosis; Sjogren's syndrome;
 KW vasculitis; sarcoidosis; autoimmune haemolytic anaemia;
 KW autoimmune thrombocytopenia; thyroiditis; diabetes mellitus;
 KW renal disease; demyelinating disease; central nervous system;
 KW peripheral nervous system; demyelinating polyneuropathy;
 KW Guillain-Barre syndrome;
 KW Chronic inflammatory demyelinating polyneuropathy.
 XX OS Homo sapiens.

XX PN WO2004043361-A2.
 XX PD 27-MAY-2004.
 XX PP 06-NOV-2003; 2003WO-US035268.
 XX PR 08-NOV-2002; 2002US-0425235P.
 XX PA (GETH) GENENTECH INC.
 XX PI Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
 PI Wood WI, Wu TD;
 XX DR WPI; 2004-420067/39.
 XX DR P-PSDB; ADO20321.
 XX PT Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
 PT treating an immune related disorder such as systemic lupus erythematosus,
 PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
 PT spondyloarthropathy.
 XX PS Claim 1; SEQ ID NO 1294; 1731pp; English.
 XX CC The invention relates to human PRO polypeptides and the polynucleotides
 CC encoding them. The polypeptides and polynucleotides are useful for
 CC treating and diagnosing immune related disorders in mammals. The immune
 CC related disorders include systemic lupus erythematosus, rheumatoid
 CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
 CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
 CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
 CC mellitus, immune-mediated renal disease, demyelinating diseases of the
 CC central or peripheral nervous system, demyelinating polyneuropathy,
 CC Guillain-Barre syndrome and chronic inflammatory demyelinating
 CC polyneuropathy. This sequence represents a human PRO polynucleotide of
 CC the invention.
 XX SQ Sequence 8570 BP; 2643 A; 1538 C; 1611 G; 2778 T; 0 U; 0 Other;
 Query Match 100.0%; Score 14; DB 12; Length 8570;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCGCCCTT 14
 DB 4603 TTGCCCCGCCCTT 4616
 RESULT 31
 ADP55596
 ID ADP55596 standard; cDNA; 8570 BP.
 XX AC ADP55596;
 XX DT 18-NOV-2004 (first entry)
 XX DE Human PRO cDNA sequence SEQ ID NO:1572.
 XX KW human; PRO; immune related disease; inflammatory immune response;
 KW immune response stimulation; anti-allergic; anti-naemic; anti-arthritis;
 KW antiasthmatic; antidiabetic; anti-inflammatory; antipsoriatic;
 KW antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;
 KW haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;
 KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
 KW virucide; Gene therapy; Gene; ss.
 XX OS Homo sapiens.
 XX PN WO2004039956-A2.
 XX PD 13-MAY-2004.
 XX PR 28-OCT-2003; 2003WO-US034381.

```
XX 29-OCT-2002; 2002US-0422472P.
XX (GETH ) GENENTECH INC.
XX Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
XX Wood WI, Wu TD;
XX WPI: 2004-376182/35.
XX P-PSDB; ADP55597.
XX New PRO polynucleotides and polypeptides, useful in useful in diagnosing
XX and treating an immune related disease, e.g. systemic lupus
XX erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in
XX stimulating an immune response.
XX Claim 2; SEQ ID NO 1572; 3009pp; English.
XX The present invention describes an isolated PRO nucleic acid (I). Also
XX described: (1) a vector comprising (1); (2) a host cell comprising the
XX vector of (1); (3) a process for producing a PRO polypeptide; (4) an
XX isolated PRO polypeptide; (5) a chimeric molecule comprising the
XX polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
XX antibody which specifically binds to a polypeptide of (4); (7) a
XX composition of matter comprising a polypeptide of (4), an agonist or
XX antagonist of the polypeptide or an antibody that binds to the
XX polypeptide in combination with a carrier; (8) an article of manufacture
XX comprising a container, a label on the container and a composition of
XX matter of (7); (9) a method of treating an immune related disease in a
XX mammal; (10) a method for determining the presence of a PRO polypeptide
XX in a sample suspected of having the polypeptide; (11) a method of
XX diagnosing an immune related disease or an inflammatory immune response
XX in a mammal; (12) a method of identifying a compound that inhibits or
XX mimics the activity of or expression of a gene encoding a PRO polypeptide
XX ; and (13) a method of stimulating the immune response in a mammal. The
XX PRO sequences have anti-allergic, antianaemic, antiarthritic,
XX antiasthmatic, antidiabetic, antiinflammatory, antipsoriatic,
XX antirheumatic, antithyroid, CNS, dermatological, gastrointestinal,
XX haemostatic, hepatoprotective, immunostimulant, immunosuppressive, muscular,
XX nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and
XX virucide activities, and can be used in gene therapy. The nucleic acid
XX (I) and the encoded polypeptides, compositions, kits and methods are
XX useful in diagnosing and treating an immune related disease and in
XX stimulating an immune response. The present sequence represents a human
XX PRO nucleotide sequence from the present invention.
XX Sequence 8570 BP; 2643 A; 1538 C; 1611 G; 2778 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 14; DB 13; Length 8570;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCCCTT 14
Db 4603 TTGCCCCGCCCCCTT 4616
RESULT 32
ABZ74489/c
ID ABZ74489 standard; DNA; 9760 BP.
XX ABZ74489;
XX 12-MAY-2003 (first entry)
XX Secreted protein gene 313 genomic fragment HT5FX79, SEQ ID NO:1636.
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
XX drug screening; chromosome identification; chromosome mapping;
XX cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
XX antianaemic; vulnery; gene; ds.
```

```
XX Homo sapiens.
XX WO200277013-A2.
XX 03-OCT-2002.
XX 26-MAR-2002; 2002WO-US009370.
XX 27-MAR-2001; 2001US-0278650P.
XX 12-SEP-2001; 2001US-00950082.
XX 12-SEP-2001; 2001US-00950083.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2003-040578/03.
XX New human secreted proteins and nucleic acids, useful for detecting or
XX treating cancer or other hyperproliferative disorders, autoimmune
XX disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX Disclosure; Page 2217-2219; 2474pp; English.
XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
XX protein genes, and ABP00947-ABP01363 represent the proteins they encode.
XX ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
XX invention also encompasses antibodies specific for the secreted proteins,
XX the use of the secreted proteins in drug screening and recombinant
XX vectors and host cells comprising a nucleic acid of the invention. The
XX secreted proteins are thought to be involved in biological activities
XX associated with cellular signalling, cellular differentiation, cell
XX migration, prohormone activation and neurotransmitter activity. The
XX secreted proteins, nucleic acids encoding them, antibodies or antibody
XX fragments specific for the secreted proteins, and modulators of protein
XX activity are useful for diagnosing or treating cancers or other
XX hyperproliferative disorders. Additionally, the secreted proteins and
XX their nucleic acids may also be used in the treatment of autoimmune
XX disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
XX (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
XX wound healing. Nucleic acids of the invention may be used for chromosome
XX identification, chromosome mapping, in gene therapy, for identifying
XX individuals from minute biological samples, as hybridisation probes, and
XX as molecular weight markers. The present sequence represents a human
XX secreted protein genomic fragment referred to in the disclosure of the
XX invention
XX Sequence 9760 BP; 2404 A; 2528 C; 2584 G; 2244 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 14; DB 8; Length 9760;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCCCTT 14
Db 8473 TTGCCCCGCCCCCTT 8460
RESULT 33
ADA44498/c
ID ADA44498 standard; DNA; 9760 BP.
XX ADA44498;
XX 20-NOV-2003 (first entry)
XX Human secreted protein DNA SEQ ID 691.
XX Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
XX Neuroprotective; Cerebroprotective; Antianemic; ds.
XX Homo sapiens.
```



```
XX WO2003000865-A2.
XX 03-JAN-2003.
XX 26-MAR-2002; 2002WO-US009105.
XX 27-MAR-2001; 2001US-0278650P.
XX 12-SEP-2001; 2001US-00950082.
XX 12-SEP-2001; 2001US-00950083.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI; 2003-184045/18.
XX A human secreted protein and nucleic acids useful for preparing a
PT diagnostic or pharmaceutical composition for diagnosing or treating
PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
PT retinopathy, neuropathy.
XX Disclosure; SEQ ID NO 691; 701pp; English.
XX The invention relates to novel genes and their fragments which are useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids and proteins
CC are useful in the diagnosis, treatment and prevention of conditions
CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
CC polyneuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
CC infection, cataract, renal disorders, or endocrine disorders. The present
CC sequence was used to illustrate the invention.
XX Sequence 9760 BP; 2404 A; 2528 C; 2584 G; 2244 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 14; DB 8; Length 9760;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
DB 8473 TTGCCCCGCCCTT 8460
RESULT 34
AAS36849
ID AAS36849 standard; DNA; 13273 BP.
XX AAS36849;
XX 17-DEC-2001 (first entry)
DE Human cardiovascular system antigen genomic DNA SEQ ID No 2349.
XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nontropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; Gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; de;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX Homo sapiens.
OS
XX WO200155321-A2.
XX 02-AUG-2001.
XX
```

```
PP 17-JAN-2001; 2001WO-US001340.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
XX 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
```


PR 19-MAY-2000; 2000US-0205515P.
PR 28-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241828P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764869.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Barash SC;

WPI; 2003-743766/70.

New cardiovascular system related polynucleotides and polypeptides, useful for preventing, treating, or ameliorating a medical condition, such as cancer of cardiovascular tissues and cancer metastases.

Claim 1; SEQ ID NO 2349; 262pp; English.

The invention relates to human cardiovascular system related polypeptides and the polynucleotides encoding them. The polypeptides, polynucleotides and antibodies to the polypeptides are useful for diagnosing a

CC pathological condition or a susceptibility to a pathological condition,
CC for preventing, treating, or ameliorating a medical condition, such as
CC cancer of cardiovascular system tissues, proliferative disorders, foetal
CC and developmental abnormalities, haematopoietic disorders, diseases of
CC the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
CC arthritis), inflammation, allergies, neurological disorders (e.g.,
CC Alzheimer's disease, Parkinson's disease), cognitive disorders,
CC schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
CC diabetes, atherosclerosis, cardiovascular disorders, angiogenic
CC disorders, kidney disorders, gastrointestinal disorders, pregnancy-
CC related disorders, endocrine disorders and infections. The nucleic acids
CC are also useful for chromosome identification, radiation hybrid mapping
CC or long-range restriction mapping. The polypeptides and polynucleotides
CC may also be used as food additives or preservatives to increase or
CC decrease storage capabilities, fat content or other nutritional
CC components. This sequence represents human cardiovascular system related
CC genomic DNA of the invention.

XX
SQ Sequence 13273 BP; 3401 A; 3266 C; 3635 G; 2971 T; 0 U; 0 Other;

Query Match 100.0%; Score 14; DB 10; Length 13273;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14

|||||
Db 9548 TTGCCCCGCCCTT 9561

RESULT 36

ADJ08961

ID ADJ08961 standard; DNA; 13273 BP.

XX AC ADJ08961;

XX DT 04-NOV-2004 (first entry)

XX DE Human cardiovascular system associated polypeptide-related DNA SeqID2349.

XX KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW breast neoplasms; liver neoplasia; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn;
KW organ transplantation; cell culture; tissue regeneration; chemotaxis;
KW food additive; preservative; cardiovascular system associated antigen;
KW nuclear factor kappaB; NFkappaB; promoter element; human; ds.

XX OS Homo sapiens.

XX PN US2004005575-A1.

XX PD 08-JAN-2004.

XX PF 26-AUG-2002; 2002US-00227577.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-019874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-0076486P.
PR 07-MAR-2002; 2002US-00091504.
(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2004-081713/08.
DR New cardiovascular system-related nucleic acid molecule, useful for
PT diagnosing, preventing or treating diseases of the cardiovascular system,
PT and in chromosome mapping, drug screening or in pharmacogenomics.
PS Disclosure; SEQ ID NO 2349; 262pp; English.
XX The invention relates to an isolated nucleic acid molecule encoding a
CC human cardiovascular system associated polypeptide (or antigens), or its
CC fragment. Also included recombinant vectors, recombinant host cells, an
CC isolated human cardiovascular system associated polypeptide (including
CC its fragment, allelic variant, species homologue or epitope), an isolated
CC antibody that binds specifically to a human cardiovascular system
CC associated polypeptide, diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or absence of a mutation in human cardiovascular system
CC associated nucleic acid and diagnosing a condition based on the presence
CC or absence of the mutation), identifying a binding partner to human

CC cardiovascular system associated polypeptides, the gene corresponding to
CC the human cardiovascular system associated cDNA sequence and identifying
CC an activity in a biological assay comprising expressing the human
CC cardiovascular system associated cDNA in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity. The human
CC cardiovascular system associated nucleic acids and polypeptides are used
CC to prevent, treat or ameliorate a medical condition (for example in
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep), for
CC example autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders, for example neoplasms of the breast or
CC liver, cardiovascular disorders, for example cardiac arrest,

Query Match 100.0%; Score 14; DB 13; Length 13273;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14

Db 9548 TTGCCCCGCCCTT 9561

RESULT 37

AAZ60888/c

ID AAZ60888 standard; DNA; 17131 BP.

XX AC AAZ60888;

DT 16-MAY-2000 (first entry)

XX DE DNA encoding a human geranylgeranyl pyrophosphate synthetase (hGGPPS).

KW Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1;

KW 1q42-1q43 locus; prostate cancer; hGGPPS; biallelic marker;

XX mevalonic biosynthetic pathway; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT exon 486..546

FT /tag= a

FT /number= 1

FT intron 547..7291

FT /tag= b

FT /number= 1

FT exon 633..826

FT /tag= c

FT /number= 1bis

FT intron 827..7191

FT /tag= d

FT /number= 1bis

FT exon 7292..7384

FT /tag= e

FT /number= 2

FT intron 7385..13759

FT /tag= f

FT /number= 2

FT exon 13760..13830

FT /tag= g

FT /number= 3

FT intron 13831..14062

FT /tag= h

FT /number= 3

FT exon 14063..15251

FT /tag= i

FT /number= 4

PN WO200005382-A2.

XX 03-FEB-2000.

XX 23-JUL-1999; 99WO-IB001353.

XX

```

PR 23-JUL-1998; 98US-0093940F.
XX (GEST ) GENSET.
XX Bouqueleret L;
XX WPI; 2000-182704/16.
DR P-PSDB; AAY68909.
XX New isolated human geranyl-geranyl pyrophosphate synthetase nucleic
PT acids, used to develop agents for the diagnosis of, e.g. pathologies
PT related to a defect in the mevalonic biosynthetic pathway.
XX
PS Claim 1; Page 72-79; 88pp; English.
XX
CC The present sequence represents a genomic sequence of human
CC geranylgeranyl pyrophosphate synthetase (hGGPPS). The sequence comprises
CC the 5' regulatory region, the exons and introns, and 3' regulatory
CC region. Two differently spliced mRNAs exist for this gene. The first
CC spliced mRNA is derived from a cDNA (AAZ60888) which comprises exons 1,
CC 2, 3 and 4. The second mRNA is derived from a cDNA (AAZ60889) which
CC comprises 1bis, 2, 3, and 4. The hGGPPS gene is located on chromosome 1,
CC at the 1q42-1q43 locus. This chromosome 1 locus has been shown to carry a
CC predisposing gene for prostate cancer. The nucleic acids encoding hGGPPS
CC can be used for screening for agents which modulate the expression of the
CC hGGPPS gene. Such agents can be used in therapeutic applications. The
CC biallelic markers associated with the hGGPPS gene can be used for the
CC diagnosis of diseases related to an alteration in the regulatory or
CC coding regions of hGGPPS, such as pathologies related to a defect in the
CC mevalonic biosynthetic pathway. The products can also be used for
CC detection, diagnosis and drug screening
XX
SQ Sequence 17131 BP; 5110 A; 3434 C; 3759 G; 4816 T; 0 U; 12 Other;

Query Match 100.0%; Score 14; DB 3; Length 17131;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGCCCCGCCCTT 14
Db 583 TTGCCCCGCCCTT 570

RESULT 38
AAT08126/c
ID AAT08126 standard; DNA; 26698 BP.
XX
AC AAT08126;
XX
DT 13-AUG-1996 (first entry)
XX
DE Mouse syndecan-1 gene.
XX
KW Syndecan-1; ectodomain; tumour; cancer; therapy; proteoglycan;
XX cell differentiation; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT intron 1..4138
FT misc_difference 1154 /tag= a
FT exon /tag= b
FT /note= "base n is not identified in the specification"
FT 4139..4444
FT /tag= c
FT /codon_start= 4378..4380
FT intron 4445..22024
FT exon /tag= d
FT 22025..23106
FT /tag= e
FT exon 23001..23482
FT /tag= g

```

```

FT intron 23107..23000
FT /tag= f
FT intron 23483..23903
FT /tag= h
FT exon 23904..24043
FT /tag= i
FT intron 24042..24250
FT /tag= j
FT exon 24251..26298
FT /tag= k
FT intron 26299..26698
FT /tag= l
XX
PN WO9534316-A1.
XX
XX 21-DEC-1995.
XX
PF 13-JUN-1995; 95WO-FI000344.
XX
PR 13-JUN-1994; 94US-00258862.
XX
PA (JALK/) JALKANEN M.
PA (MALI/) MALI M.
XX
PI Jalkanen M, Mali M;
XX
DR WPI; 1996-049416/05.
DR P-PSDB; AAR87001.
XX
PT New syndecan ectodomain compsn. to reduce tumour growth - for delivery to
PT extracellular environment for suppressing tumour growth in malignant and
PT non-malignant tumours.
XX
PS Disclosure; Fig 2; 49pp; English.
XX
CC A mouse gene (AAT08126) codes for syndecan-1 (AAR87001), a cell surface
CC proteoglycan. The ectodomain of syndecan-1 (amino acids 18-251) is
CC capable of restoring a more differentiated morphology to tumour cells and
CC to suppress the growth of malignant cells. The DNA can be used for the
CC prodn. of recombinant ectodomain, pref. using host cells of a similar
CC cell type as that of the tumour. The human syndecan-1 gene (AAT08125) has
CC also been identified
XX
SQ Sequence 26698 BP; 5736 A; 6620 C; 7176 G; 7165 T; 0 U; 1 Other;

Query Match 100.0%; Score 14; DB 2; Length 26698;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGCCCCGCCCTT 14
Db 4027 TTGCCCCGCCCTT 4014

RESULT 39
AAQ67902/c
ID AAQ67902 standard; DNA; 26700 BP.
XX
AC AAQ67902;
XX
DT 25-MAR-2003 (revised)
DT 08-DEC-1994 (first entry)
XX
DE Syndecan gene.
XX
KW Syndecan; enhancer; differentiation; tumor; therapy; hair growth; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT intron 1..4377
FT exon /tag= a
FT 4378..4443

```

```

FT intron /*tag= b
FT 4444..22025
FT /*tag= c
FT 22026..22106
FT /*tag= d
FT 22107..23000
FT /*tag= e
FT 23001..23483
FT /*tag= f
FT 23484..23904
FT /*tag= g
FT 23905..24039
FT /*tag= h
FT 24040..24250
FT /*tag= i
FT 24251..24418
FT /*tag= j
FT 24419..26700
FT /*tag= k
XX WO9412162-A1.
PN
XX
XX 09-JUN-1994.
XX
XX 01-DEC-1993; 93WO-F1000514.
XX
XX 01-DEC-1992; 92US-00988427.
XX
XX (WAER/) WAERRI A. M.
XX (ALAN/) ALANEN-KURKI L. M.
XX (AUVI/) AUVINEN P. O. V.
XX (JAAK/) JAAKKOLA P. M.
XX (JALK/) JALKANEN M. T.
XX (LEPP/) LEPPAE S. M.
XX (MALI/) MALI M. S.
XX (VIHI/) VIHINEN T. A.
XX
XX Waerri AM, Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalkanen MT;
PI Leppae SM, Mali MS, Viuhinen TA;
XX
XX WPI; 1994-19926/24.
XX P-PSDB; AAR55276.
XX
XX Syndecan stimulation of cellular differentiation - useful for decreasing
XX tumour growth used to promote hair growth.
XX
XX Disclosure; Page 22-39; 65pp; English.
XX
XX The mouse syndecan gene enhancer, located 8-10 kb upstream from the
XX initiation site, is given in AAG67901. Manipulation of the enhancer can
XX be used either to slow or prevent tumor growth or to promote
XX differentiation of specific cell types, e.g. epidermal cells to promote
XX hair formation. The complete mouse syndecan gene and its encoded protein
XX are given in AAG67902 and AAR55276. (Updated on 25-MAR-2003 to correct PN
XX field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-
XX 2003 to correct PI field.)
XX
XX Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 0 U; 1 Other;
XX
XX Query Match 100.0%; Score 14; DB 2; Length 26700;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+03;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TTGCCCGCCGCTT 14
XX |||||
XX 4027 TTGCCCGCCGCTT 4014
XX
XX RESULT 40
XX AAV15946/c
XX ID AAV15946 standard; DNA; 26700 BP.
XX
XX AC AAV15946;

```

```

XX 28-MAY-1998 (first entry)
XX Mouse syndecan gene sequence.
XX Syndecan; tumour suppression; tissue regeneration; enhancement; mouse;
XX wound healing; ds.
XX Mus sp.
XX
XX Key Location/Qualifiers
XX intron 1..4377
XX /*tag= b
XX /number= 1
XX 4378..24421
XX /*tag= a
XX /product= "syndecan protein"
XX /note= "contains introns"
XX 4378..4443
XX /*tag= c
XX /number= 1
XX 4444..22025
XX /*tag= d
XX /number= 2
XX 22026..22106
XX /*tag= e
XX /number= 2
XX 22107..23000
XX /*tag= f
XX /number= 3
XX 23001..23483
XX /*tag= g
XX /number= 3
XX 23484..23904
XX /*tag= h
XX /number= 4
XX 23905..24039
XX /*tag= i
XX /number= 4
XX 24040..24250
XX /*tag= j
XX /number= 5
XX 24251..24418
XX /*tag= k
XX /number= 5
XX 24422..26700
XX /*tag= l
XX /number= 6
XX
XX US5726058-A.
XX 10-MAR-1998.
XX
XX 07-JUN-1995; 95US-00472217.
XX
XX 01-DEC-1992; 92US-00988427.
XX 01-DEC-1993; 93WO-F1000514.
XX 07-MAR-1994; 94US-00206186.
XX
XX (AUVI/) AUVINEN P.
XX (MALI/) MALI M.
XX (VIHI/) VIHINEN T.
XX (WAER/) WAERRI A.
XX (JALK/) JALKANEN M.
XX (ALAN/) ALANEN-KURKI L.
XX (LEPP/) LEPPAE S.
XX (JAAK/) JAAKKOLA P.
XX
XX Jaakkola P, Waerri A, Leppae S, Mali M, Alanen-Kurki L;
XX Auvinen P, Jalkanen M, Viuhinen T;
XX
XX WPI; 1998-192770/17.
XX P-PSDB; AAW47156.

```

XX New mouse syndecan gene sequences - useful for, e.g. suppressing tumour
 PT growth or promoting tissue regeneration in processes such as wound
 PT healing.
 PT
 XX Claim 2; Fig 2A-O; 48pp; English.
 XX
 CC This is the mouse syndecan gene sequence. A 350 base pair fragment
 CC (AAV15948) of a purified 2196 base pair DNA molecule (AAV15947) enhances
 CC the expression of a gene operably linked to the promoter of the mouse
 CC syndecan gene in 3T3 cells following treatment with TGF- beta and bFGF
 CC when the fragment is operably linked to the promoter. A purified DNA
 CC molecule comprising a portion of the nucleotide residues 3538-3888 of the
 CC mouse syndecan genomic sequence suppresses expression of a gene operably
 CC linked to the promoter of the mouse syndecan gene in S115 cells treated
 CC with testosterone. Host cells can be transfected with vectors which
 CC contain either the enhancing or suppressing DNA molecules. The products
 CC may be used to alter the differentiated state of a host cell by altering
 CC its expression of syndecan, e.g. to induce and regulate syndecan
 CC expression, especially in cells which exhibit a malignant phenotype,
 CC regardless of the origin of transformation. The products can be used to
 CC produce therapeutics for suppressing tumour growth. They may enhance the
 CC syndecan expression in a host cell, by enhancing its gene transcription,
 CC especially in malignant or normal cells, and therefore promote tissue
 CC regeneration, especially in processes such as wound healing
 XX
 XX Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 0 U; 1 Other;

Query Match 100.0%; Score 14; DB 2; Length 26700;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 4027 TTGCCCCGCCCTT 4014

RESULT 41
 AAV81283/c
 ID AAV81283 standard; DNA; 26700 BP.

XX AAV81283;
 XX
 XX 11-MAR-1999 (first entry)
 XX
 XX Mouse syndecan-1 encoding DNA.

XX Syndecan-1; tumour; ectodomain; epithelial; mesenchymal; breast; mouse;
 KW endometrial tumour; prostatic tumour; oestrogenic; androgenic; steroid;
 KW glioma; myeloma; carcinoma; sarcoma; lymphoma; adenoma; ss.

OS Mus sp.

XX Key Location/Qualifiers
 FH 4378..24421
 FT CDS /tag= a
 FT /product= "mouse syndecan-1"
 FT /note= "contains introns"
 FT 4378..4443
 FT /tag= b
 FT /number= 1
 FT 4444..22025
 FT /tag= c
 FT /number= 1
 FT 22026..22107
 FT /tag= d
 FT /number= 2
 FT 22108..23001
 FT /tag= e
 FT /number= 2
 FT 23002..23483
 FT /tag= f
 FT /number= 3

FT intron 23484..23904
 FT /tag= g
 FT /number= 3
 FT 23905..24040
 FT /tag= h
 FT /number= 4
 FT 24041..24251
 FT /tag= i
 FT /number= 4
 FT 24252..2418
 FT /tag= j
 FT /number= 5

PN US5851993-A.

PD 22-DEC-1998.

PF 07-JUN-1995; 95US-00488199.

PR 13-JUN-1994; 94US-00258862.

PA (BIOT-) BIOTIE THERAPIES LTD.

XX Mali M, Jalkanen M;

XX WPI; 1999-104635/09.

DR P-PSDB; AAW95198.

XX Reducing growth of tumour cells - with ectodomain of syndecan applied to
 PT the extracellular environment to induce a more differentiated phenotype,
 PT particularly for hormone-dependent breast, endometrial or prostatic
 PT cancers.

PS Disclosure; Fig 2A-M; 48pp; English.

XX The invention relates to a method of reducing the growth of tumour cells
 CC which comprises supplying a human syndecan ectodomain to the environment
 CC around the cells causing them to develop a more differentiated phenotype.
 CC The method is used to suppress tumour cells of epithelial, mesenchymal,
 CC pre-B or plasma cell origin, especially breast, endometrial or prostatic
 CC tumours, and particularly those responsive to oestrogenic or non-
 CC steroid. More generally it can be used to treat malignant or non-
 CC malignant tumours, particularly those characterised by loss of syndecan,
 CC e.g. gliomas, myelomas, carcinomas, sarcomas, lymphomas and adenomas. The
 CC present sequence represents the DNA sequence encoding a mouse syndecan-1
 XX Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 0 U; 1 Other;

Query Match 100.0%; Score 14; DB 2; Length 26700;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14

|||||

Db 4027 TTGCCCCGCCCTT 4014

RESULT 42

ACA60750/c

ID ACA60750 standard; DNA; 26700 BP.

XX ACA60750;

XX 27-JUN-2003 (first entry)

XX Mouse cell surface proteoglycan, Syndecan-1, gene.

XX Mouse; ds; gene; Syndecan-1; cutaneous wound healing; hair growth;
 KW tissue regeneration; syndecan regulation; syndecan expression;
 KW cell surface proteoglycan.

XX Mus sp.

OS

PH Key Location/Qualifiers
FT CDS 4378..24421
FT /tag= a
FT /product= "Syndecan-1"
FT exon 4378..4443
FT /tag= b
FT /number= 1
FT intron 4444..22025
FT /tag= c
FT /number= 1
FT exon 22026..22106
FT /tag= b
FT /number= 2
FT intron 22107..23000
FT /tag= c
FT /number= 2
FT /cons splice= (5'site.No,3'site.No)
FT exon 23001..23483
FT /tag= b
FT intron 23484..23904
FT /tag= c
FT exon 23905..24040
FT /tag= b
FT /number= 4
FT intron 24041..24250
FT /tag= c
FT /number= 4
FT /cons splice= (5'site.No,3'site.No)
FT exon 24251..24421
FT /tag= b
FT /number= 5
US6492344-B1.
PN
XX
PD 10-DEC-2002.
XX
XX
XX 21-JUN-1999; 99US-00336757.
XX
PR 01-DEC-1993; 93WO-FI000514.
PR 07-MAR-1994; 94US-00206186.
XX
PA (BIOT-) BIOTIE THERAPIES CORP.
XX
XX Jalkanen M, Jaakkola P, Vihinen T;
XX WPI; 2003-391071/37.
DR P-PSDB; AB009486.
XX
XX Promotion of cutaneous wound healing comprises introducing to wound site recombinant expression vector comprising syndecan enhancer element linked to promoter which is linked to structural gene encoding growth factor.
PT
PT
XX
XX Disclosure; Fig 2; 84pp; English.
XX
XX The invention relates to a method of promoting cutaneous wound healing which comprises directly introducing to a cutaneous wound site a recombinant expression vector comprising the syndecan enhancer element, FGF-inducible Response Element (FIRE), operably linked to a FIRE-activated promoter which is operably linked to a structural gene encoding a growth factor that promotes the cutaneous wound healing. The method is for promoting regeneration of tissue e.g. promotion of hair growth. The invention enhances syndecan expression via administration of growth factors. The enhancer element regulates the expression of syndecan. The present sequence represents the mouse cell surface proteoglycan, Syndecan -1, gene
XX
SQ Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 0 U; 1 Other;
Query Match 100.0%; Score 14; DB 8; Length 26700;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
DB 4027 TTGCCCCCGCCCTT 4014
RESULT 43
ADJ10262/c
ID ADJ10262 standard; DNA; 51001 BP.
XX
AC ADJ10262;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human geranylgeranyl diphosphate synthase 1 genomic DNA SeqID 11.
XX
KW gene; da; chromosome 1q43; human; geranylgeranyl diphosphate synthase 1; GGPS1; geranylgeranyl pyrophosphate synthetase; GGPPS; gpppsase;
KW geranyltransferase; embryonic development; cell differentiation;
KW apoptosis; developmental; hyperproliferative disorder; cancer;
KW cytostatic.
OS Homo sapiens.
XX
XX US2004005570-A1.
XX
XX 08-JAN-2004.
XX
XX 02-JUL-2002; 2002US-00189268.
XX
XX 02-JUL-2002; 2002US-00189268.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Dean NM, Bennett CF, Dobie KW;
XX WPI; 2004-201146/19.
XX
XX New antisense oligonucleotides for modulating geranylgeranyl diphosphate synthase 1 expression, useful for diagnosing, preventing or treating conditions associated with the protein, e.g. cancer.
PS Example 15; SEQ ID NO 11; 76pp; English.
XX
XX This invention relates to a novel antisense compounds that modulate the expression of human geranylgeranyl diphosphate synthase 1 (also known as GGPS1, geranylgeranyl pyrophosphate synthetase, GGPPS, gpppsase and geranyltransferase) and located on chromosome 1p43. Specifically, it refers to compositions useful for inhibiting the expression of GGPS1, which normally participates in embryonic development, cell differentiation and stimulates apoptosis via caspase-3 activation. The present invention describes antisense oligonucleotides that comprise at least one modified sugar moiety, a 2'-O-methoxyethyl (2' MOE) and at least one modified nucleobase, a 5-methylcytosine. Accordingly, these compounds are useful for treating a disease or condition associated with GGPS1 such as a developmental or hyperproliferative disorder (e.g. cancer) that arise as a result of aberrant apoptosis. As such, these compositions exhibit cytostatic activity and are useful for diagnostics, prophylaxis, research reagents and various kits. This polynucleotide sequence is the geranylgeranyl diphosphate synthase 1 genomic DNA of the invention.
XX
SQ Sequence 51001 BP; 13906 A; 11464 C; 11661 G; 13770 T; 0 U; 200 Other;
Query Match 100.0%; Score 14; DB 12; Length 51001;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
DB 1912 TTGCCCCCGCCCTT 1899

```

RESULT 44
AAA71425/c
ID AAA71425 standard; DNA; 27 BP.
XX
AC AAA71425;
XX
DT 01-DEC-2000 (first entry)
XX
DE Murine pathogenic prion protein inhibiting peptide PCR primer #2.
XX
KW Antiprion; antineurodegenerative; cellular prion protein inhibitor;
KW pathological isoform; diagnostic marker; fatal familial insomnia; FFI;
KW transmissible spongiform encephalopathy; BSE; Huntington's disease;
KW new variant Creutzfeld-Jakob disease; bovine spongiform encephalopathy;
KW Gerstmann-Strauszler-Scheinker disease; Alzheimer's disease; detection;
KW neurodegenerative disease; murine; PCR primer; ss.
XX
OS Mus sp.
XX
PN DE19917838-A1.
XX
PD 13-JUL-2000.
XX
PF 20-APR-1999; 99DE-01017838.
XX
PR 12-JAN-1999; 99DE-01000860.
XX
PA (BUND ) BUNDESREPUBLIK DEUT PAUL-EHRLICH-INST.
XX
PI Schroeder B, Caughey B, Mong C, Kurth R;
XX
WPI; 2000-499915/45.
XX
PT Composition for treating prion or other neurodegenerative diseases,
PT comprises a peptide that inhibits unfolding of the prion to its
PT pathological isoform.
XX
PS Example 3; Page 15; 18pp; German.
XX
CC This invention describes a novel pharmaceutical composition comprising an
CC 18-mer peptide (I) or its modified and/or concatameric forms, or analogs
CC (collectively (II)) which have antiprion and antineurodegenerative
CC activity. (I) inhibit conversion of cellular prion protein to the
CC pathological isoform. (I), also nucleic acid that encodes them, are used:
CC (i) as diagnostic markers of transmissible spongiform encephalopathies
CC (specifically (new variant) Creutzfeld-Jakob disease, Gerstmann-
CC Strausler-Scheinker disease, fatal familial insomnia (FFI), bovine
CC spongiform encephalopathy (BSE)), Alzheimer's and Huntington's diseases
CC and/or other neurodegenerative diseases; and (ii) for prevention and or
CC treatment of these diseases, in humans or other animals. Antibodies
CC against (I) are used as diagnostic reagents for detecting (I). This
CC sequence represents a PCR primer used in the amplification of the murine
CC pathogenic prion protein inhibiting peptide encoding DNA described in the
CC method of the invention
XX
SQ Sequence 27 BP; 6 A; 4 C; 16 G; 1 T; 0 U; 0 Other;

Query Match 92.9%; Score 13; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCCCTT 14
DB 16 TGCCCCGCCCCCTT 4

RESULT 45
ABK98227
ID ABK98227 standard; DNA; 34 BP.
XX
AC ABK98227;
XX
DT 07-OCT-2002 (first entry)
XX

```

```

XX Triple helix forming associated DNA sequence #54.
DE
XX Triple-helix formation; purine-rich target sequence; double-helix DNA;
KW gene expression; regulatory sequence; pathogenic double-stranded DNA;
KW pathogenic bacteria; virus; replication; virulence; cancer;
KW oncogene suppression; cancerous cell; cytostatic; antimicrobial; ss.
XX
OS Synthetic.
XX
PN US6403302-B1.
XX
PD 11-JUN-2002.
XX
PF 16-DEC-1993; 93US-00168920.
XX
PR 17-SEP-1992; 92US-00946976.
XX
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Dervan PB, Beal PA;
XX
WPI; 2002-536030/57.
XX
PT A triple-helix comprising a double helical nucleic acid (DHNA) and an
PT oligonucleotide which binds in parallel and antiparallel orientation,
PT respectively, for targeting sequences on alternate strands of DHNA to
PT control gene expression.
XX
PS Example 9; Fig 37A; 108pp; English.
XX
CC The present invention relates to methods and oligonucleotides for forming
CC a triple-helix comprising a double helical nucleic acid comprising first
CC and second substantially complementary strands, and an oligonucleotide
CC bound to a purine-rich target sequence within the double helical nucleic
CC acid, where the oligonucleotide binds in a parallel and antiparallel
CC orientation, respectively, to target sequences on alternate strands of
CC the double helical nucleic acid. The method has therapeutic applications,
CC where gene expression is controlled by selective triple-helix formation
CC within expression regulatory sequences of a target gene. The
CC oligonucleotides can be used to form triple-helices, and are useful to
CC detect the presence or absence of specific sequences within genomic DNA
CC for diagnostic and therapeutic purposes. The oligonucleotides can be
CC selected to specifically bind to pathogenic double-stranded DNA including
CC specific sequences required by pathogenic bacteria or viruses for
CC replication or virulence, reducing their pathogenicity. Alternatively,
CC the oligonucleotide can be chosen to target a unique sequence of the
CC pathogen which is not found in the genome of pathogen's host. The
CC oligonucleotides can be used in cancer treatment by way of triple-helix
CC suppression of specific oncogenes including those of endogenous or viral
CC origin. Such therapeutic oligonucleotides are capable of forming triple-
CC helices with such sequences in cancerous cells containing the activated
CC oncogene, so preferentially killing or repressing the cancer causing
CC cell. The present sequence represents a DNA sequence used in the methods
CC of the present invention
XX
SQ Sequence 34 BP; 3 A; 14 C; 4 G; 13 T; 0 U; 0 Other;

Query Match 92.9%; Score 13; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCCCTT 14
DB 2 TGCCCCGCCCCCTT 14

RESULT 46
ABK98226/c
ID ABK98226 standard; DNA; 35 BP.
XX
AC ABK98226;
XX

```

```

DT 07-OCT-2002 (first entry)
DE Triple helix forming associated DNA sequence #53.
XX Triple-helix formation; purine-rich target sequence; double-helix DNA;
KW gene expression; regulatory sequence; pathogenic double-stranded DNA;
KW pathogenic bacteria; virus; replication; virulence; cancer;
KW oncogene suppression; cancerous cell; cytostatic; antimicrobial; ss.
OS Synthetic.
XX
XX US6403302-B1.
XX
XX 11-JUN-2002.
XX
XX 16-DEC-1993; 93US-00168920.
XX
XX 17-SEP-1992; 92US-00946976.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Dervan PB, Beal PA;
XX
XX WPI; 2002-536030/57.
XX
XX A triple-helix comprising a double helical nucleic acid (DHNA) and an
PT oligonucleotide which binds in parallel and antiparallel orientation,
PT respectively, for targeting sequences on alternate strands of DHNA to
PT control gene expression.
XX
XX Example 9; Fig 37A; 108pp; English.
XX
XX The present invention relates to methods and oligonucleotides for forming
CC a triple-helix comprising a double helical nucleic acid comprising first
CC and second substantially complementary strands, and an oligonucleotide
CC bound to a purine-rich target sequence within the double helical nucleic
CC acid, where the oligonucleotide binds in a parallel and antiparallel
CC orientation, respectively, to target sequences on alternate strands of
CC the double helical nucleic acid. The method has therapeutic applications,
CC where gene expression is controlled by selective triple-helix formation
CC within expression regulatory sequences of a target gene. The
CC oligonucleotides can be used to form triple-helices, and are useful to
CC detect the presence or absence of specific sequences within genomic DNA
CC for diagnostic and therapeutic purposes. The oligonucleotides can be
CC selected to specifically bind to pathogenic double-stranded DNA including
CC specific sequences required by pathogenic bacteria or viruses for
CC replication or virulence, reducing their pathogenicity. Alternatively,
CC the oligonucleotide can be chosen to target a unique sequence of the
CC pathogen which is not found in the genome of pathogen's host. The
CC oligonucleotides can be used in cancer treatment by way of triple-helix
CC suppression of specific oncogenes including those of endogenous or viral
CC origin. Such therapeutic oligonucleotides are capable of forming triple-
CC helices with such sequences in cancerous cells containing the activated
CC oncogene, so preferentially killing or repressing the cancer causing
CC cell. The present sequence represents a DNA sequence used in the methods
CC of the present invention
XX
XX Sequence 35 BP; 13 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
XX
XX Query Match 92.9%; Score 13; DB 6; Length 35;
XX Best Local Similarity 100.0%; Pred. No. 5.1e+03;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 TGCCTCCGCCCTT 14
XX | | | | |
XX Db 34 TGCCTCCGCCCTT 22
XX
XX RESULT 47
XX ABK98129/c
XX ID ABK98129 standard; DNA; 40 BP.
XX
XX AC ABK98129;

DT 07-OCT-2002 (first entry)
DE Triple helix forming associated DNA sequence #53.
XX Triple-helix formation; purine-rich target sequence; double-helix DNA;
KW gene expression; regulatory sequence; pathogenic double-stranded DNA;
KW pathogenic bacteria; virus; replication; virulence; cancer;
KW oncogene suppression; cancerous cell; cytostatic; antimicrobial; ss.
OS Synthetic.
XX
XX US6403302-B1.
XX
XX 11-JUN-2002.
XX
XX 16-DEC-1993; 93US-00168920.
XX
XX 17-SEP-1992; 92US-00946976.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Dervan PB, Beal PA;
XX
XX WPI; 2002-536030/57.
XX
XX A triple-helix comprising a double helical nucleic acid (DHNA) and an
PT oligonucleotide which binds in parallel and antiparallel orientation,
PT respectively, for targeting sequences on alternate strands of DHNA to
PT control gene expression.
XX
XX Example 9; Fig 37A; 108pp; English.
XX
XX The present invention relates to methods and oligonucleotides for forming
CC a triple-helix comprising a double helical nucleic acid comprising first
CC and second substantially complementary strands, and an oligonucleotide
CC bound to a purine-rich target sequence within the double helical nucleic
CC acid, where the oligonucleotide binds in a parallel and antiparallel
CC orientation, respectively, to target sequences on alternate strands of
CC the double helical nucleic acid. The method has therapeutic applications,
CC where gene expression is controlled by selective triple-helix formation
CC within expression regulatory sequences of a target gene. The
CC oligonucleotides can be used to form triple-helices, and are useful to
CC detect the presence or absence of specific sequences within genomic DNA
CC for diagnostic and therapeutic purposes. The oligonucleotides can be
CC selected to specifically bind to pathogenic double-stranded DNA including
CC specific sequences required by pathogenic bacteria or viruses for
CC replication or virulence, reducing their pathogenicity. Alternatively,
CC the oligonucleotide can be chosen to target a unique sequence of the
CC pathogen which is not found in the genome of pathogen's host. The
CC oligonucleotides can be used in cancer treatment by way of triple-helix
CC suppression of specific oncogenes including those of endogenous or viral
CC origin. Such therapeutic oligonucleotides are capable of forming triple-
CC helices with such sequences in cancerous cells containing the activated
CC oncogene, so preferentially killing or repressing the cancer causing
CC cell. The present sequence represents a DNA sequence used in the methods
CC of the present invention
XX
XX Sequence 35 BP; 13 A; 5 C; 14 G; 3 T; 0 U; 0 Other;
XX
XX Query Match 92.9%; Score 13; DB 6; Length 35;
XX Best Local Similarity 100.0%; Pred. No. 5.1e+03;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 TGCCTCCGCCCTT 14
XX | | | | |
XX Db 34 TGCCTCCGCCCTT 22
XX
XX RESULT 48
XX ADS47155/c
XX ID ADS47155 standard; cDNA; 207 BP.
XX
XX AC ADS47155 standard; cDNA; 207 BP.

```

AC ADS47155;
XX
XX
XX 02-DEC-2004 (first entry)
XX Bacterial polynucleotide #1898.
DE
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
XX Bacteria.
XX
XX US2003233675-A1.
PN
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
PF
XX 21-FEB-2002; 2002US-0360039P.
PR
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 25585; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 207 BP; 87 A; 33 C; 55 G; 32 T; 0 U; 0 Other;
SQ
Query Match 92.9%; Score 13; DB 13; Length 207;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 TCCCCCGCCCTT 14
|||||

Db 194 TCCCCCGCCCTT 182
RESULT 49
AAD43323/c
ID AAD43323 standard; DNA; 238 BP.
XX
XX AAD43323;
XX
XX 14-NOV-2002 (first entry)
DT
DE Human DTC DNA fragment comprising promoter and partial exon 1.
XX
XX Human; tumour associated protein; DTC protein; cancer; immune response;
KW gene therapy; vaccine; cytostatic; chromosome 4p11; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH promoter 1..215
FT exon /tag= a
FT 216..238
FT /tag= b
FT /number= 1
FT /partial
XX
XX US2002081608-A1.
PN
XX 27-JUN-2002.
XX
XX 19-JUL-2001; 2001US-00909147.
XX
XX 21-JUL-2000; 2000US-0219674P.
PR
XX 05-MAR-2001; 2001US-0273451P.
XX
XX (LUOL/) LUO L.
PA (DIAM/) DIAMANDIS E P.
XX
XX Luo L, Diamandis EP;
PI WPI; 2002-635463/68.
XX
XX New isolated nucleic acid molecules encoding tumor associated proteins,
CC useful for diagnosing, monitoring, preventing or treating cancer,
CC particularly ovarian and testicular tumors, by triggering immune
CC responses in patients.
XX
XX Example; Fig 5; 42pp; English.
XX
XX The present invention relates to novel polynucleotides encoding tumour
CC associated proteins, designated DTC proteins. The nucleic acid molecules
CC are useful in diagnosing, monitoring, preventing or treating cancer. They
CC trigger immune responses in cancer patients. The compounds and antibodies
CC are useful for treating disorders involving tumour-associated proteins.
CC Sequences of the invention are useful in gene therapy. They are also used
CC as vaccines. The present sequence is human DTC DNA fragment comprising
CC promoter and partial exon 1. DTC gene is located on chromosome 4p11
XX
XX Sequence 238 BP; 38 A; 76 C; 73 G; 51 T; 0 U; 0 Other;
SQ
Query Match 92.9%; Score 13; DB 6; Length 238;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 TCCCCCGCCCTT 14
|||||
Db 52 TCCCCCGCCCTT 40
RESULT 50
ADC21271
ID ADC21271 standard; DNA; 260 BP.
XX

AC ADC21271;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human cell-specific expression vector-related DNA sequence #2.
XX
KW human; cell-specific expression vector; cell-specific replication vector;
KW gene therapy; malignant tumour; liver fibrosis; postoperative stenosis;
KW organ transplantation; arteriosclerosis; diabetic omentopathy; ds.
XX
OS Homo sapiens.
XX
PN WO2003057888-A1.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-JP013683.
XX
PR 28-DEC-2001; 2001JP-00402102.
PR 30-AUG-2002; 2002JP-00255395.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Takahashi K, Yamamura H;
XX
DR WPI; 2003-577527/54.
XX
PT Cell-specific expression/replication vector containing transcription
PT initiation regulating domain of human calponin gene, applicable in gene
PT therapy of malignant tumor lung and liver fibrosis or diabetic
PT omentopathy.
XX
PS Claim 3; SEQ ID NO 2; 66pp; Japanese.
XX
CC The invention comprises a cell-specific expression/replication vector
CC which does not act on normal cells. The vector of the invention is useful
CC in gene therapy for the treatment of: malignant tumour, liver fibrosis,
CC postoperative stenosis (e.g. stenosis after organ transplantation),
CC arteriosclerosis, and diabetic omentopathy. The present DNA sequence
CC represents a human sequence of the invention.
XX
SQ Sequence 260 BP; 61 A; 82 C; 82 G; 35 T; 0 U; 0 Other;

Query Match 92.9%; Score 13; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCGCGCCCTT 14
DB 84 TGCCCGCGCCCTT 96

Search completed: April 25, 2005, 11:36:55
Job time : 155.638 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:13:24 ; Search time 1204.72 Seconds
(without alignments)
442.342 Million cell updates/sec

Title: US-10-010-476-15

Perfect score: 14

Sequence: 1 TTGCCCCGCCCTT 14

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : EST:

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gsa1:*

9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	55	1	AA896461
2	14	100.0	83	5	BQ129142
3	14	100.0	110	2	AW141226
4	14	100.0	113	5	BQ127833
5	14	100.0	122	2	BE863436
6	14	100.0	132	2	AW902893
7	14	100.0	136	7	CR383072
8	14	100.0	154	7	CK291373
9	14	100.0	161	6	CB721578
10	14	100.0	164	2	BE076558
11	14	100.0	166	4	BG949054
12	14	100.0	175	1	AA430712
13	14	100.0	191	6	CB721279
14	14	100.0	198	9	CE643460
15	14	100.0	200	8	AQ843933
16	14	100.0	223	5	BU755470
17	14	100.0	224	7	CN114354
18	14	100.0	240	3	CNS08WOU
19	14	100.0	242	1	AV212144
20	14	100.0	247	7	CR383085
21	14	100.0	250	8	BH019647
22	14	100.0	255	1	AI909047
23	14	100.0	265	1	AA200152
24	14	100.0	267	2	BF710971

C	25	14	100.0	267	4	BI451845
	26	14	100.0	283	7	CF094505
	27	14	100.0	286	7	CN863008
	28	14	100.0	286	9	CV150513
	29	14	100.0	288	9	CC807244
	30	14	100.0	292	4	BG358747
	31	14	100.0	296	7	CR383121
	32	14	100.0	300	2	BF261295
	33	14	100.0	300	7	W14598
	34	14	100.0	314	1	AA136032
	35	14	100.0	314	5	BU028163
	36	14	100.0	317	1	AA033451
	37	14	100.0	327	7	CN863792
	38	14	100.0	327	7	R00632
	39	14	100.0	329	5	BU215685
	40	14	100.0	331	1	AA030175
	41	14	100.0	335	2	BB543943
	42	14	100.0	336	4	BG941842
	43	14	100.0	338	9	CC852229
	44	14	100.0	345	7	CV318910
	45	14	100.0	345	8	AQ106680
	46	14	100.0	350	4	BG939965
	47	14	100.0	350	5	BQ343488
	48	14	100.0	351	4	BG277108
	49	14	100.0	352	7	CR382632
	50	14	100.0	353	7	CR382747
	51	14	100.0	353	7	W08476
	52	14	100.0	357	1	AA427004
	53	14	100.0	357	5	BQ243399
	54	14	100.0	358	7	W53193
	55	14	100.0	361	4	BG939966
	56	14	100.0	364	7	R07778
	57	14	100.0	364	7	R46503
	58	14	100.0	365	7	T73524
	59	14	100.0	367	2	BF745493
	60	14	100.0	372	1	AA919985
	61	14	100.0	372	2	BF600530
	62	14	100.0	373	1	A1267052
	63	14	100.0	375	2	BF429978
	64	14	100.0	375	8	BH019682
	65	14	100.0	376	1	AA137538
	66	14	100.0	376	5	BU605354
	67	14	100.0	377	7	T49247
	68	14	100.0	377	7	CF615180
	69	14	100.0	379	1	AL633254
	70	14	100.0	379	7	CV318867
	71	14	100.0	381	4	BG671781
	72	14	100.0	382	2	BF736952
	73	14	100.0	384	4	BG023360
	74	14	100.0	385	4	BI939207
	75	14	100.0	385	7	CN892543
	76	14	100.0	386	2	BF658054
	77	14	100.0	387	1	AA032587
	78	14	100.0	387	7	CF751352
	79	14	100.0	388	1	AA638828
	80	14	100.0	388	1	AA215005
	81	14	100.0	388	5	BY038935
	82	14	100.0	391	2	BB755190
	83	14	100.0	394	1	AA874568
	84	14	100.0	394	5	BY424326
	85	14	100.0	394	6	BY659869
	86	14	100.0	397	2	BB734364
	87	14	100.0	398	8	BH039837
	88	14	100.0	399	9	CE424900
	89	14	100.0	402	2	BB811591
	90	14	100.0	404	2	BF430075
	91	14	100.0	405	5	BP426975
	92	14	100.0	405	5	BY425967
	93	14	100.0	407	1	AA821806
	94	14	100.0	407	1	AA821807
	95	14	100.0	407	1	AA089272
	96	14	100.0	407	1	BF048307
	97	14	100.0	408	6	BY652852

BI451845	TY18504.h
CF094505	QNR18C01.
CN863008	000916AAL
CV150513	MGst6009k
CC807244	ZMMBCC048
BG358747	BOVMS1-00
CR383121	CR383121
BF261295	IM0007H12
W14598	mb28C01.r1
AA136032	zn88608.r
BU028163	QHH1OK10.
AA033451	mi42806.r
CN863792	000919AAL
R00632	ye74906.r1
BU215685	603106847
AA030175	mh86C01.r
BB543943	BB543943
BG941842	ax18805.x
CC852229	NDL.101M1
CV318910	CM2-KT003
AQ106680	HS.3080.B
BG939965	ax01G08.x
BQ343488	IL3-NT010
BG277108	ux63e08.y
CR382632	CR382632
CR382747	CR382747
W08476	mb46a06.r1
AA427004	VG78B05.r
BQ243399	TaE15015E
W53193	md41f07.r1
BG939966	ax01G08.y
R07778	yf15b07.r1
R46503	yj52a09.r1
T73524	YC35f01.r1
BF745493	CM2-BT082
AA919985	yy22G07.r
BF600530	265249.MA
A1267052	UJ10C02.x
BF429978	255916.MA
BH019682	L460k.d.H
AA137538	mg77G04.r
BU605354	mah9C09.
T49247	ya73f03.r1
CF615180	CES09113
AL633254	AL633254
CV318867	CM2-IT003
BG671781	DRNBXH04
BB736952	BB736952
BG023360	Gab62e11.
BI939207	Gab7Bh06.
CN892543	010603AAx
BF658054	masa94e09.
AA032587	ml32Q03.r
CF751352	UI-M-HK0-
AA638828	vn73G06.r
AA215005	mu83d07.r
BY038935	BY038935
BB755190	BB755190
AA874568	vw84h10.r
BY424326	BY424326
BY659869	BY659869
BB734364	BB734364
BH039837	RFC1-24-2
CE424900	tigr-gss-
BB811591	BB811591
BF430075	256025.WA
BP426975	BP426975
BY425967	BY425967
AA821806	vp72C05.r
AA821807	vp72C07.r
AA089272	mc60a02.x
BF048307	db81a06.y
BY652852	BY652852

98	14	100.0	408	7	CF614897	CF614897 CES008559	c 171	14	100.0	511	9	CR143204	CR143204 Reverse s
99	14	100.0	413	1	AV663245	AV663245 VM663245	c 172	14	100.0	514	1	AI743763	AI743763 wg53f04.x
100	14	100.0	413	1	AA571883	AA571883 VM64f08.r	c 173	14	100.0	514	4	BM069090	BM069090 id73d10.x
101	14	100.0	414	2	BB788953	BB788953 BB788953	c 174	14	100.0	515	5	BQ129025	BQ129025 i126d11.x
102	14	100.0	420	1	AU232680	AU232680 AU232680	c 175	14	100.0	515	7	CK775125	CK775125 BE93939.MA
103	14	100.0	420	1	AU232835	AU232835 AU232835	c 176	14	100.0	516	7	CM440305	CM440305 BE904021B2
104	14	100.0	421	7	CN940634	CN940634 U10913AVB	c 177	14	100.0	516	8	BZ263208	BZ263208 CH230-375
105	14	100.0	421	8	BZ118210	BZ118210 CH230-317	c 178	14	100.0	518	6	CB681481	CB681481 OSJNEF07M
106	14	100.0	425	2	BB680308	BB680308 BB680308	c 179	14	100.0	523	4	BM740300	BM740300 K-EST0011
107	14	100.0	427	2	BB675591	BB675591 BB675591	c 180	14	100.0	524	7	CK345082	CK345082 10041A10-
108	14	100.0	430	5	BY437918	BY437918 BY437918	c 181	14	100.0	526	4	BI465839	BI465839 1e20h11.x
109	14	100.0	431	1	AA387177	AA387177 vc18a09.r	c 182	14	100.0	526	7	CK892300	CK892300 SGPI52319
110	14	100.0	432	1	AA182364	AA182364 mt77c03.r	c 183	14	100.0	532	1	AA219990	AA219990 mv65g06.r
111	14	100.0	432	5	BQ044156	BQ044156 UI-M-EPO-	c 184	14	100.0	532	2	BE134544	BE134544 ug16c02.y
112	14	100.0	433	1	AA561753	AA561753 vk9a11.r	c 185	14	100.0	533	6	CA601876	CA601876 wr1.pk001
113	14	100.0	433	2	BE948385	BE948385 UI-M-BH3	c 186	14	100.0	537	6	CB122494	CB122494 K-EST0170
114	14	100.0	434	2	BE136228	BE136228 ug28a07.x	c 187	14	100.0	538	8	AQ848129	AQ848129 LMAJFV1.1
115	14	100.0	436	1	AA242567	AA242567 zw46e10.r	c 188	14	100.0	539	4	BI976871	BI976871 486034.MA
116	14	100.0	436	2	BF430395	BF430395 933.MARC	c 189	14	100.0	541	5	BQ354249	BQ354249 PM3-HT090
117	14	100.0	437	2	BE098278	BE098278 UI-R-BJ1-	c 190	14	100.0	544	2	BB730755	BB730755 BB730755
118	14	100.0	437	5	BQ51219	BQ51219 i173f08.x	c 191	14	100.0	544	5	BQ587198	BQ587198 E012349-0
119	14	100.0	437	7	CK389828	CK389828 L0948F05-	c 192	14	100.0	545	5	BQ748183	BQ748183 UI-M-FAO-
120	14	100.0	438	7	CN906569	CN906569 o11001ABD	c 193	14	100.0	545	5	BX684891	BX684891 BX684891
121	14	100.0	439	1	AA049002	AA049002 mj45c12.r	c 194	14	100.0	545	5	BY477127	BY477127 BY477127
122	14	100.0	439	5	BQ337849	BQ337849 QV3-NN102	c 195	14	100.0	547	4	BI889168	BI889168 ZF637-2-0
123	14	100.0	439	5	BY453911	BY453911 BY453911	c 196	14	100.0	547	7	CK483001	CK483001 AGENCOURT
124	14	100.0	440	6	CB749944	CB749944 AMGNNUC-N	c 197	14	100.0	550	7	CK950134	CK950134 4075098.B
125	14	100.0	441	2	BE101015	BE101015 UI-R-BJ1-	c 198	14	100.0	551	6	CB128967	CB128967 K-EST0178
126	14	100.0	445	2	BB674104	BB674104 BB674104	c 199	14	100.0	555	4	BM125474	BM125474 id74b12.x
127	14	100.0	446	1	AA839088	AA839088 vm47d09.r	c 200	14	100.0	555	5	BQ748111	BQ748111 UI-M-FAO-
128	14	100.0	447	1	AA914280	AA914280 vy99a11.r	c 201	14	100.0	559	9	CC849073	CC849073 NDL.123M3
129	14	100.0	447	2	BE985921	BE985921 UI-M-CG0P	c 202	14	100.0	560	1	A1414947	A1414947 mc82d02.x
130	14	100.0	448	5	BQ128974	BQ128974 i125d04.x	c 203	14	100.0	560	7	CF614549	CF614549 CES008775
131	14	100.0	450	1	AA798997	AA798997 vv95h02.r	c 204	14	100.0	561	6	C79806	C79806 C79806.Mous
132	14	100.0	450	1	AV593443	AV593443 AV593443	c 205	14	100.0	565	7	CO094796	CO094796 GR_Ea17B
133	14	100.0	450	2	BE135185	BE135185 ug28a07.y	c 206	14	100.0	573	8	BH294057	BH294057 CH230-122
134	14	100.0	453	7	H68968	H68968 yr85f06.r1	c 207	14	100.0	574	5	BQ385148	BQ385148 NISC.mn10
135	14	100.0	456	5	BY451138	BY451138 BY451138	c 208	14	100.0	581	2	BE784737	BE784737 601473644
136	14	100.0	458	2	BF429673	BF429673 1453.MARC	c 209	14	100.0	583	7	CO075154	CO075154 GR_Ea35K
137	14	100.0	459	4	BI791107	BI791107 id05h05.x	c 210	14	100.0	584	6	CR0147	CR0147 C80147.Mous
138	14	100.0	459	5	BY452023	BY452023 BY452023	c 211	14	100.0	590	1	AA543168	AA543168 vJ81g11.r
139	14	100.0	460	9	CC571176	CC571176 CH240.446	c 212	14	100.0	590	7	CO717427	CO717427 DG14-4a24
140	14	100.0	465	2	BB832216	BB832216 BB832216	c 213	14	100.0	593	1	AV740078	AV740078 IMV0504C05
141	14	100.0	466	1	AA416018	AA416018 vf40c03.r	c 214	14	100.0	595	2	AZ641772	AZ641772 1M0504C05
142	14	100.0	472	6	BY556647	BY556647 BY556647	c 215	14	100.0	595	2	BE229232	BE229232 98B50357
143	14	100.0	474	1	AJ597600	AJ597600 AJ597600	c 216	14	100.0	597	8	BH050350	BH050350 RPCI-24-3
144	14	100.0	474	1	AV592844	AV592844 AV592844	c 217	14	100.0	598	2	AW322375	AW322375 uo74g09.x
145	14	100.0	475	4	BM823885	BM823885 K-EST0095	c 218	14	100.0	600	4	BI989827	BI989827 H3045B08-
146	14	100.0	475	6	CB337498	CB337498 laa34f09.	c 219	14	100.0	600	4	BM687757	BM687757 UI-B-CRO-
147	14	100.0	477	1	A1366138	A1366138 ao94g05.x	c 220	14	100.0	601	4	CA064112	CA064112 ssa1rqb53
148	14	100.0	478	7	W53563	W53563 md15e03.r1	c 221	14	100.0	601	6	AL897276	AL897276 AL897276
149	14	100.0	480	4	BG066544	BG066544 H3045B08-	c 222	14	100.0	603	1	BI077943	BI077943 602871972
150	14	100.0	480	4	BG223836	BG223836 1M0020B07	c 223	14	100.0	606	4	CF978548	CF978548 CF978548
151	14	100.0	480	5	BX520766	BX520766 BX520766	c 224	14	100.0	606	7	BE101111	BE101111 UI-R-BJ1-
152	14	100.0	483	2	AW476437	AW476437 ug75g10.y	c 225	14	100.0	608	2	BQ390534	BQ390534 NISC.mn13
153	14	100.0	483	2	BB862375	BB862375 UI-M-AJ1-	c 226	14	100.0	610	5	CO713104	CO713104 DG14-1118
154	14	100.0	484	1	AJ593085	AJ593085 AJ593085	c 227	14	100.0	611	8	BH038607	BH038607 RPCI-24-3
155	14	100.0	485	4	BM105542	BM105542 509020.MA	c 228	14	100.0	612	1	AL958763	AL958763 AL958763
156	14	100.0	488	4	BM429875	BM429875 1duo22F10	c 229	14	100.0	613	5	BQ391130	BQ391130 NISC.mn17
157	14	100.0	489	1	A1551948	A1551948 vi47b10.y	c 230	14	100.0	613	5	CV083099	CV083099 Mds6f003b
158	14	100.0	490	6	CA688558	CA688558 wlm96.pk0	c 231	14	100.0	616	1	AV740963	AV740963 BX740963
159	14	100.0	493	2	BF147760	BF147760 us41a06.x	c 232	14	100.0	618	5	BX693168	BX693168 BX693168
160	14	100.0	493	5	BG381289	BG381289 LM24HW014	c 233	14	100.0	618	7	CN354456	CN354456 170006001
161	14	100.0	496	4	BI429147	BI429147 fr73d06.y	c 234	14	100.0	620	8	BH019721	BH019721 L472k.d.H
162	14	100.0	496	4	BM791912	BM791912 K-EST0072	c 235	14	100.0	621	7	CO683320	CO683320 DG11-167H
163	14	100.0	496	8	AQ902037	AQ902037 LMAJFV1.1	c 236	14	100.0	623	5	BQ127641	BQ127641 i126d11.y
164	14	100.0	500	7	CV374568	CV374568 PM3-HT090	c 237	14	100.0	623	5	CN485078	CN485078 hx22g10.y
165	14	100.0	501	4	BG840755	BG840755 MEST11-E0	c 238	14	100.0	626	4	BG335097	BG335097 602403575
166	14	100.0	502	7	W45987	W45987 mc81f05.r1	c 239	14	100.0	629	1	AL972064	AL972064 AL972064
167	14	100.0	502	8	AQ800689	AQ800689 HS5315.B	c 240	14	100.0	629	5	BX696729	BX696729 BX696729
168	14	100.0	505	1	AA241769	AA241769 mw26h10.r	c 241	14	100.0	630	1	AL647301	AL647301 AL647301
169	14	100.0	509	4	BI847444	BI847444 469277.MA	c 242	14	100.0	630	5	BQ385485	BQ385485 NISC.mn12
170	14	100.0	509	4	BI847444	BI847444 469277.MA	c 243	14	100.0	630	5	BQ385485	BQ385485 NISC.mn12

244	14	100.0	630	7	CF614901	CF614901 CES008586	C 317	14	100.0	723	4	BI765180	BI765180 603051434
245	14	100.0	632	2	BB629172	BB629172 BB629172	C 318	14	100.0	723	4	BI820053	BI820053 603037238
246	14	100.0	633	4	BG681653	BG681653 602627958	C 319	14	100.0	723	5	BK753278	BK753278 BX753278
247	14	100.0	633	4	BG744469	BG744469 602723021	C 320	14	100.0	724	4	BI413686	BI413686 602986927
248	14	100.0	633	6	CA069179	CA069179 SC5BAD105	C 321	14	100.0	724	5	BI119223	BI119223 BX119223
249	14	100.0	636	7	CO086856	CO086856 GR_Ea05A	C 322	14	100.0	724	6	CB493708	CB493708 omvktcbs5
250	14	100.0	637	7	CN879591	CN879591 010419AAS	C 323	14	100.0	724	7	CK943396	CK943396 4067414 B
251	14	100.0	637	7	CO716741	CO716741 DG14-2e5	C 324	14	100.0	725	4	BI888520	BI888520 ZF637-2-0
252	14	100.0	640	9	CR236002	CR236002 Reverse s	C 325	14	100.0	726	7	CF520006	CF520006 AGENCOURT
253	14	100.0	642	7	CF762637	CF762637 CES007570	C 326	14	100.0	730	7	CK945227	CK945227 4069849 B
254	14	100.0	643	7	BN354455	BN354455 170005315	C 327	14	100.0	731	8	CC154188	CC154188 CSU-K34.1
255	14	100.0	644	5	BM88941	BM88941 LM24HW003	C 328	14	100.0	733	7	CO396234	CO396234 AGENCOURT
256	14	100.0	647	7	CN119134	CN119134 EC2CA9DB	C 329	14	100.0	734	7	CN790716	CN790716 4125339 B
257	14	100.0	647	7	CO043881	CO043881 UI-M-FW0-	C 330	14	100.0	736	2	BE786675	BE786675 601475025
258	14	100.0	650	7	CF613872	CF613872 CES007856	C 331	14	100.0	737	4	BG111881	BG111881 602283493
259	14	100.0	651	1	AL801908	AL801908 AL801908	C 332	14	100.0	737	9	AG173209	AG173209 Pan trogl
260	14	100.0	651	2	BE257132	BE257132 601108875	C 333	14	100.0	738	2	BE393877	BE393877 601312165
261	14	100.0	651	6	CD214678	CD214678 P9M2n.pX0	C 334	14	100.0	738	4	BG473622	BG473622 602515974
262	14	100.0	652	2	BF128843	BF128843 601811028	C 335	14	100.0	738	4	BG489467	BG489467 602517843
263	14	100.0	652	6	CA318263	CA318263 UI-M-FW0-	C 336	14	100.0	738	5	BH860102	BH860102 AGENCOURT
264	14	100.0	652	9	CR108786	CR108786 Reverse s	C 337	14	100.0	738	8	AZ955372	AZ955372 2M0221M23
265	14	100.0	653	5	BQ08476	BQ08476 1030004G0	C 338	14	100.0	741	4	BG024879	BG024879 602275793
266	14	100.0	655	6	CA081158	CA081158 SCACAM204	C 339	14	100.0	743	5	BK750038	BK750038 BX750038
267	14	100.0	655	7	CO089153	CO089153 GR_Ea081	C 340	14	100.0	745	2	BE270458	BE270458 601184489
268	14	100.0	655	8	AZ992208	AZ992208 2M0276G16	C 341	14	100.0	747	2	BE281131	BE281131 601157113
269	14	100.0	658	4	BI284848	BI284848 UI-R-CX0s	C 342	14	100.0	747	4	BI088759	BI088759 602851548
270	14	100.0	659	7	CO087059	CO087059 GR_Ea05F	C 343	14	100.0	747	6	CA250448	CA250448 SCRUFL111
271	14	100.0	663	2	BB667105	BB667105 BB667105	C 344	14	100.0	748	9	AG496138	AG496138 Mus muscu
272	14	100.0	665	6	CD623425	CD623425 56044631J	C 345	14	100.0	750	9	CR212010	CR212010 Reverse s
273	14	100.0	667	6	CD623424	CD623424 56044631H	C 346	14	100.0	751	4	BG763861	BG763861 602736814
274	14	100.0	668	2	BF306783	BF306783 601891765	C 347	14	100.0	752	7	CO592490	CO592490 D62-62012
275	14	100.0	668	4	BG705039	BG705039 602686554	C 348	14	100.0	753	7	CV121415	CV121415 AGENCOURT
276	14	100.0	668	4	BM767961	BM767961 K-EST0050	C 349	14	100.0	755	4	BI086436	BI086436 602849890
277	14	100.0	669	2	BF686577	BF686577 602143491	C 350	14	100.0	756	9	CR276729	CR276729 Reverse s
278	14	100.0	670	7	CF614743	CF614743 CES009147	C 351	14	100.0	758	2	BE562473	BE562473 601335950
279	14	100.0	672	6	CA239864	CA239864 SCMCFL802	C 352	14	100.0	761	5	BK756132	BK756132 BX756132
280	14	100.0	672	8	CO067133	CO067133 CSU-K33R	C 353	14	100.0	761	6	CB207713	CB207713 AGENCOURT
281	14	100.0	674	1	AV362172	AV362172 AGENCOURT	C 354	14	100.0	761	7	CO402949	CO402949 AGENCOURT
282	14	100.0	675	6	CA463161	CA463161 AGENCOURT	C 355	14	100.0	763	5	BK755141	BK755141 BX755141
283	14	100.0	676	5	BU655392	BU655392 1112119F0	C 356	14	100.0	763	7	CO884369	CO884369 BovGen.12
284	14	100.0	678	6	BY759646	BY759646 BX759646	C 357	14	100.0	766	7	CF270383	CF270383 AGENCOURT
285	14	100.0	678	6	CA051007	CA051007 ssaargb52	C 358	14	100.0	766	9	CC847129	CC847129 ND.L.96019
286	14	100.0	679	6	BG706558	BG706558 602672931	C 359	14	100.0	767	4	BG481196	BG481196 602528519
287	14	100.0	679	6	BY707029	BY707029 BY707029	C 360	14	100.0	767	4	BI550798	BI550798 603195694
288	14	100.0	680	5	BK684574	BK684574 BX684574	C 361	14	100.0	771	9	AG581112	AG581112 Mus muscu
289	14	100.0	682	2	BB121838	BB121838 BB121838	C 362	14	100.0	772	8	AO621837	AO621837 HS 3107 B
290	14	100.0	682	5	BP695271	BP695271 BP695271	C 363	14	100.0	773	4	BI456335	BI456335 603172842
291	14	100.0	683	7	CN294917	CN294917 170006003	C 364	14	100.0	773	7	CF218233	CF218233 AGENCOURT
292	14	100.0	686	2	BE314517	BE314517 601148521	C 365	14	100.0	775	7	CF150402	CF150402 AGENCOURT
293	14	100.0	686	5	BP765462	BP765462 BP765462	C 366	14	100.0	784	4	BG744626	BG744626 602722716
294	14	100.0	687	2	BE795656	BE795656 601590641	C 367	14	100.0	785	4	BG828029	BG828029 602752083
295	14	100.0	687	4	BG751689	BG751689 602730411	C 368	14	100.0	786	2	BE544324	BE544324 601076812
296	14	100.0	687	6	CB169180	CB169180 VBB603020	C 369	14	100.0	788	5	BG761960	BG761960 602718911
297	14	100.0	688	7	CO398984	CO398984 AGENCOURT	C 370	14	100.0	788	5	EX683510	EX683510 BX683510
298	14	100.0	689	4	BG281477	BG281477 602402003	C 371	14	100.0	788	7	CF217693	CF217693 AGENCOURT
299	14	100.0	692	5	BQ175309	BQ175309 UI-M-DJ2-	C 372	14	100.0	788	7	CF222547	CF222547 AGENCOURT
300	14	100.0	693	4	BQ066728	BQ066728 H3046C06-	C 373	14	100.0	788	7	CK774797	CK774797 963588 NA
301	14	100.0	698	5	BQ207154	BQ207154 UI-R-D21-	C 374	14	100.0	792	7	CF150396	CF150396 AGENCOURT
302	14	100.0	700	4	BI415984	BI415984 602987106	C 375	14	100.0	794	4	BG696984	BG696984 602659070
303	14	100.0	701	5	BM945726	BM945726 UI-M-EMO-	C 376	14	100.0	794	5	BK754017	BK754017 BX754017
304	14	100.0	705	2	BE887462	BE887462 601508108	C 377	14	100.0	795	4	BI414295	BI414295 602986586
305	14	100.0	706	7	CN354451	CN354451 170006001	C 378	14	100.0	796	4	BI412299	BI412299 602987526
306	14	100.0	708	4	BG743038	BG743038 602632161	C 379	14	100.0	796	4	BI905076	BI905076 603169409
307	14	100.0	709	5	BY773652	BY773652 BX773652	C 380	14	100.0	797	2	BE274981	BE274981 601122514
308	14	100.0	710	4	BI093093	BI093093 602858196	C 381	14	100.0	797	9	CR256529	CR256529 Reverse s
309	14	100.0	712	6	CB321157	CB321157 AGENCOURT	C 382	14	100.0	798	4	BI528559	BI528559 602925410
310	14	100.0	714	5	BQ176178	BQ176178 UI-M-DJ2-	C 383	14	100.0	801	9	CR268369	CR268369 Reverse s
311	14	100.0	714	7	CN354447	CN354447 170004240	C 384	14	100.0	803	4	BI247561	BI247561 602960014
312	14	100.0	714	9	CC528985	CC528985 CH240 404	C 385	14	100.0	804	2	BK758823	BK758823 602125511
313	14	100.0	716	1	AL650763	AL650763 AL650763	C 386	14	100.0	804	2	BG758823	BG758823 602713185
314	14	100.0	719	1	CF121303	CF121303 UI-RF-CB0	C 387	14	100.0	805	2	BE277152	BE277152 601178685
315	14	100.0	720	4	BG334511	BG334511 602461144	C 388	14	100.0	805	7	CF243026	CF243026 AGENCOURT
316	14	100.0	722	4	BI761204	BI761204 603043435	C 389	14	100.0	806	7	CF217348	CF217348 AGENCOURT

C 390	14	100.0	808	4	BI553471	BI553471	603193234	C 463	14	100.0	874	4	BG699894	BG699894	602681248
C 391	14	100.0	808	5	BX730190	BX730190	CSU-K33r.	C 464	14	100.0	874	4	BI553890	BI553890	603190850
C 392	14	100.0	808	8	CO082015	CO082015	CSU-K33r.	C 465	14	100.0	874	5	BX752457	BX752457	603190850
C 393	14	100.0	809	5	BX868925	BX868925	BX868925	C 466	14	100.0	874	5	BX759593	BX759593	603190850
C 394	14	100.0	810	4	BG863118	BG863118	602651602	C 467	14	100.0	875	5	BUI83474	BUI83474	AGENCYCOURT
C 395	14	100.0	811	5	BX866301	BX866301	602651602	C 468	14	100.0	875	7	CF272129	CF272129	AGENCYCOURT
C 396	14	100.0	811	5	BX872990	BX872990	602651602	C 469	14	100.0	875	7	CF272129	CF272129	AGENCYCOURT
C 397	14	100.0	812	5	BX778279	BX778279	602651602	C 470	14	100.0	878	5	BX384603	BX384603	603046775
C 398	14	100.0	812	6	CB990776	CB990776	AGENCYCOURT	C 471	14	100.0	880	5	BU553288	BU553288	AGENCYCOURT
C 399	14	100.0	814	4	BG581078	BG581078	602628905	C 472	14	100.0	880	5	BX701061	BX701061	AGENCYCOURT
C 400	14	100.0	814	9	AG101681	AG101681	Pan trogl	C 473	14	100.0	881	5	BG622657	BG622657	602647413
C 401	14	100.0	815	4	BI414796	BI414796	602990711	C 474	14	100.0	883	5	BQ425247	BQ425247	AGENCYCOURT
C 402	14	100.0	815	5	BU744594	BU744594	CH1#001.D	C 475	14	100.0	884	5	BX764534	BX764534	AGENCYCOURT
C 403	14	100.0	815	7	CK652830	CK652830	AGENCYCOURT	C 476	14	100.0	885	2	BF127965	BF127965	601810715
C 404	14	100.0	816	2	BF676759	BF676759	602086413	C 477	14	100.0	885	4	BG748446	BG748446	602706418
C 405	14	100.0	816	4	BI904893	BI904893	603168929	C 478	14	100.0	886	4	BI903703	BI903703	603167577
C 406	14	100.0	816	4	BM041208	BM041208	603614393	C 479	14	100.0	886	5	BUI55806	BUI55806	AGENCYCOURT
C 407	14	100.0	816	5	BU596686	BU596686	AGENCYCOURT	C 480	14	100.0	887	5	BX383869	BX383869	AGENCYCOURT
C 408	14	100.0	817	2	BE780565	BE780565	601468871	C 481	14	100.0	887	5	BF760272	BF760272	603046775
C 409	14	100.0	817	5	BU744564	BU744564	CH1#001.B	C 482	14	100.0	887	7	CF882089	CF882089	603046775
C 410	14	100.0	820	3	BC051595	BC051595	Mus muscu	C 483	14	100.0	888	5	BQ572338	BQ572338	AGENCYCOURT
C 411	14	100.0	822	4	BG482562	BG482562	602529306	C 484	14	100.0	889	5	BU956783	BU956783	AGENCYCOURT
C 412	14	100.0	825	7	CF219534	CF219534	AGENCYCOURT	C 485	14	100.0	889	5	BX762713	BX762713	603167577
C 413	14	100.0	828	9	CR224129	CR224129	Forward s	C 486	14	100.0	889	9	BX971169	BX971169	Forward s
C 414	14	100.0	830	5	BX882656	BX882656	60262656	C 487	14	100.0	889	6	CB235486	CB235486	AGENCYCOURT
C 415	14	100.0	830	7	BN824584	BN824584	Oa splbn	C 488	14	100.0	890	9	CG233783	CG233783	OG3AU59TV
C 416	14	100.0	834	4	BG281186	BG281186	602402111	C 489	14	100.0	891	5	BQ948019	BQ948019	AGENCYCOURT
C 417	14	100.0	834	4	BG757754	BG757754	602714863	C 490	14	100.0	891	9	CR147899	CR147899	Reverse s
C 418	14	100.0	836	4	BI903526	BI903526	603168083	C 491	14	100.0	892	4	BG181040	BG181040	602329145
C 419	14	100.0	837	4	BI413191	BI413191	602990086	C 492	14	100.0	893	5	BQ877366	BQ877366	AGENCYCOURT
C 420	14	100.0	837	5	BU912718	BU912718	AGENCYCOURT	C 493	14	100.0	894	5	BQ437933	BQ437933	AGENCYCOURT
C 421	14	100.0	838	4	BG751572	BG751572	602730679	C 494	14	100.0	894	6	CD514264	CD514264	AGENCYCOURT
C 422	14	100.0	838	6	CD653489	CD653489	AGENCYCOURT	C 495	14	100.0	895	4	BI412859	BI412859	602988176
C 423	14	100.0	838	8	CC445457	CC445457	PUHZ52TB	C 496	14	100.0	895	5	BQ233188	BQ233188	AGENCYCOURT
C 424	14	100.0	839	4	BI094428	BI094428	602856261	C 497	14	100.0	895	8	CC116617	CC116617	NDL.94K4.
C 425	14	100.0	839	4	BI820679	BI820679	603034464	C 498	14	100.0	896	2	BF698725	BF698725	602126263
C 426	14	100.0	839	9	CNS049RZ	CNS049RZ	Tetraodon	C 499	14	100.0	896	5	BQ918207	BQ918207	AGENCYCOURT
C 427	14	100.0	840	4	BG331661	BG331661	602432655	C 500	14	100.0	899	5	BQ943599	BQ943599	AGENCYCOURT
C 428	14	100.0	841	2	BF214591	BF214591	601845906								
C 429	14	100.0	841	4	BI525297	BI525297	602924568								
C 430	14	100.0	841	4	BI820012	BI820012	603035386								
C 431	14	100.0	843	6	CA988635	CA988635	AGENCYCOURT								
C 432	14	100.0	843	8	CC118826	CC118826	NDL.31N10								
C 433	14	100.0	846	2	BF664502	BF664502	602146260								
C 434	14	100.0	846	7	CF219593	CF219593	AGENCYCOURT								
C 435	14	100.0	846	9	CNS03854	AL239953	Tetraodon								
C 436	14	100.0	847	7	CK773393	CK773393	962025 MA								
C 437	14	100.0	848	4	BG705050	BG705050	602686570								
C 438	14	100.0	850	4	BG705050	BG705050	602686570								
C 439	14	100.0	852	5	BX847251	BX847251	602686570								
C 440	14	100.0	854	1	AU117859	AU117859	60117859								
C 441	14	100.0	854	4	BI546520	BI546520	603191531								
C 442	14	100.0	857	2	BE778769	BE778769	601465108								
C 443	14	100.0	857	4	BI729665	BI729665	603349344								
C 444	14	100.0	858	4	BG336590	BG336590	602405695								
C 445	14	100.0	858	9	CNS007UX	AL050867	Drosophila								
C 446	14	100.0	859	2	BE959216	BE959216	601645313								
C 447	14	100.0	861	7	CO331270	CO331270	AGENCYCOURT								
C 448	14	100.0	862	5	BQ049339	BQ049339	AGENCYCOURT								
C 449	14	100.0	862	9	CL666453	CL666453	PR10152c								
C 450	14	100.0	864	2	BF534917	BF534917	602050860								
C 451	14	100.0	864	5	BUI17160	BUI17160	AGENCYCOURT								
C 452	14	100.0	866	5	BU959191	BU959191	AGENCYCOURT								
C 453	14	100.0	866	8	CC081079	CC081079	CSU-K33r.								
C 454	14	100.0	867	7	CNS23489	CNS23489	Oa splbn								
C 455	14	100.0	868	2	BE791912	BE791912	601585808								
C 456	14	100.0	869	2	BF026427	BF026427	601671448								
C 457	14	100.0	869	4	BI414453	BI414453	602986068								
C 458	14	100.0	870	2	BE564934	BE564934	601343646								
C 459	14	100.0	870	5	BQ232675	BQ232675	AGENCYCOURT								
C 460	14	100.0	870	5	BUI181505	BUI181505	AGENCYCOURT								
C 461	14	100.0	871	9	CL069344	CL069344	CH216-117								
C 462	14	100.0	873	4	BG745042	BG745042	602723134								

ALIGNMENTS

RESULT 1	AA896461	55 bp	mRNA	linear	EST 06-APR-1998
LOCUS	AA896461	vx3b07.rl	Stratagene mouse macrophage (#937306)	Mus musculus	CDNA
DEFINITION	AA896461	clone IMAGE:1279861	5', mRNA sequence.		
ACCESSION	AA896461				
VERSION	AA896461.1	GI:3032854			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 55)				
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.				
TITLE	The WashU-HMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:671661				

```

/dev_stages=embryonic day E16.5
/lab_host="DH10B"
/clone_lib="Melton Amplified Mouse E16 5 Pancreas 3 M16S1

```

musculus cDNA clone IMAGE:5942979 5', mRNA sequence.

ACCESSION
BQ127833
VERSION
BQ127833.1 GI:20201744
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM

REFERENCE
1 (bases 1 to 113)
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Muridae; Murinae; Mus.
AUTHORS
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blisstein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y., and Bowers, Y.

TITLE
Endocrine Pancreas Consortium
COMMENT
Unpublished (2000)

JOURNAL
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center This clone is
available royalty-free through LML; please contact the IMAGE
consortium (info@image.llnl.gov) for further information
Seq primer: -40RP from Gibco.

FEATURES
source

1..113
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5942979"
/sex="Both"
/tissue_type="pancreas"
/dev_stage="Embryonic day E16.5"
/lab_host="DH10B"
/clone_lib="Melton Amplified Mouse E16 5 Pancreas 3 M16S1 A"

/note="Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 0.97 kb. Amplified once on solid support. cDNA library Preparation: Guolin Chen."

ORIGIN

Query Match 100.0%; Score 14; DB 5; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCCCTT 14
|||||
Db 52 TTGCCCCGCCCCCTT 65

RESULT 5
BE863436/c
LOCUS
DEFINITION
UI-M-AJ1-ahf-e-02-0-UI r2 NIH BMAP MOB N Mus musculus cDNA clone
UI-M-AJ1-ahf-e-02-0-UI 5', mRNA sequence.
EST.
ACCESSION
BE863436
VERSION
BE863436.1 GI:10383414
KEYWORDS
EST.

SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 122)
Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov

CDNA Library Preparation: M.B. Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.

FEATURES
source

1..122
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AJ1-ahf-e-02-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH BMAP MOB N"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The NIH BMAP MOB N library is a normalized library constructed from mouse olfactory bulbs. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

ORIGIN

Query Match 100.0%; Score 14; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCCCTT 14
|||||
Db 26 TTGCCCCGCCCCCTT 13

RESULT 6

AW902893
LOCUS
QV3-NN1025-100500-183-h05 NN1025 Homo sapiens cDNA, mRNA sequence.
DEFINITION
AW902893
VERSION
AW902893.1 GI:8067098
KEYWORDS
EST.

SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 132)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE

JOURNAL
MEDLINE
PUBMED
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV3-NN1025-100
500-183-h05&t3=2000-05-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 132.

FEATURES

source

1. .132
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1025"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from QRESNES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 100.0%; Score 14; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
Db 101 TTGCCCCGCCCTT 114

RESULT 7

CR383072/c
LOCUS CR383072 136 bp mRNA linear EST 04-MAY-2004
DEFINITION CR383072 Bovine multi-stage muscles library (bcas) Bos taurus cDNA
clone bcas0002a.a.08 5prim, mRNA sequence.
CR383072
ACCESSION CR383072.1 GI:47004976
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM

REFERENCE

AUTHORS Chevalet, C.
TITLE AGENAE, a French Animal Genome project
JOURNAL Unpublished (2004)
COMMENT Contact: Martin P
INRA, Genomique & Physiologie de la Lactation
Domaine de Vilvert, 78352 JOUY-EN-JOSAS cedex, FRANCE
Tel: +33 (0) 1.34.65.25.82
Fax: +33 (0) 1.34.65.29.26
Email: Patrice.Martine@jouy.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at eigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.

FEATURES

source

0002 row: a column: 8.
Location/Qualifiers
1. .136
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="bcas0002a.a.08"
/tissue_type="muscles : heart, longissimus thoracis,

semiteindinosus, masseter, cutaneous trunci"
/dev_stage="from embryos to adults"
/clone_lib="Bovine multi-stage muscles library (bcas)"
/note="Clone distribution : AGENAE Resource Centre.
Francois PUMI, Francois.Pumi@jouy.inra.fr, INRA, CEA
Radiobiologie et Etude du genome (LREG), Domaine de
Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0)
1.34.65.28.02, +33 (0) 1.34.65.22.73"

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
Db 86 TTGCCCCGCCCTT 73

RESULT 8

CK291373 154 bp mRNA linear EST 02-AUG-2004
LOCUS EST754087 Nicotiana benthamiana mixed tissue cDNA library.
DEFINITION normalised, full-length Nicotiana benthamiana cDNA clone NBMC173 5'
end, mRNA sequence.
ACCESSION CK291373
VERSION CK291373.1 GI:39871750
KEYWORDS EST.
SOURCE Nicotiana benthamiana
ORGANISM

REFERENCE

AUTHORS Buell, C.R., Hart, A., Zismann, V., Karameycheva, S.A., Day, B.,
Stakawicz, B., Jin, H. and Baker, B.
TITLE Generation of EST sequences from Nicotiana benthamiana
JOURNAL Unpublished (2003)
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

source

1. .154
Location/Qualifiers
/organism="Nicotiana benthamiana"
/mol_type="mRNA"
/db_xref="taxon:4100"
/clone="NBMC173"
/tissue_type="abiotic and biotic stress-treated leaves,
callus tissue and root tissue"
/lab_host="DH10B-ronA"
/clone_lib="Nicotiana benthamiana mixed tissue cDNA
library, normalized, full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Nicotiana benthamiana
tissues that include callus, roots from liquid culture
grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
challenged leaves (Pseudomonas syringae pv tomato 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
campestris pv vesicatoria 18hr). RNA was isolated from
these tissues and pooled in approximately equal molar
amounts."

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 TTGCCCCCGCCCTT 14
Db 33 TTGCCCCCGCCCTT 46

RESULT 9
LOCUS CB721578/c
DEFINITION jnn612B02.f op Ornithodoros porcinus porcinus cDNA 5', mRNA
ACCESSION CB721578
VERSION 1
KEYWORDS EST.
SOURCE Ornithodoros porcinus porcinus
ORGANISM Ornithodoros porcinus porcinus
REFERENCE Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
AUTHORS Parasitiformes; Ixodida; Ixodidae; Argasidae; Ornithodoros.
1 (bases 1 to 161)
Neilan,J.G., Kutish,G.F., Lu,Z., Burrage,T.G., Zsak,A. and
Rock,D.L.
TITLE Sequence analysis of Ornithodoros porcinus whole tick cDNA
libraries
JOURNAL Unpublished (2003)
COMMENT Contact: Neilan JG
Plum Island Animal Disease Center
US Department of Agriculture, Agricultural Research Service
PO Box 848, Greenport, NY 11944-848, USA
Tel: 631 323 3133
Fax: 631 323 3044
Email: jneilan@piadc.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim alt option. Vector identified by
cross match v0.990329 and Lucy v1.17p.
Seq primer: M13 Forward
FEATURES
source
1..161
/organism="Ornithodoros porcinus porcinus"
/mol_type="mRNA"
/sub_species="porcinus"
/db_xref="taxon:34596"
/tissue_type="Whole Tick"
/lab_host="DH10B"
/clone_lib="op"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;
Library made from polyA selected RNA, isolated from whole
stage N2 ticks two weeks after swine blood meal."

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCCGCCCTT 14
Db 60 TTGCCCCCGCCCTT 47

RESULT 10
LOCUS BE076558/c
DEFINITION CMI-BT0596-060200-116-b08 BT0596 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE076558
VERSION BE076558.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
```

```

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2=CM1-BT0596-060
200-116-b08&t3=2000-02-06&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 164.
FEATURES
source
1..164
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BT0596"
/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

ORIGIN
Query Match 100.0%; Score 14; DB 2; Length 164;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCCGCCCTT 14
Db 137 TTGCCCCCGCCCTT 124

RESULT 11
LOCUS BG949054
DEFINITION CM2-HB0003-020101-664-h06 HB0003 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG949054
VERSION BG949054.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
```

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ci=CM2&t2=CM2-HB0003-020101-664-h06&t3=2001-01-02&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 18

High quality sequence stop: 166.

Location/Qualifiers

1..166

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HB0003"

/note="Organ: bocio tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 14; DB 4; Length 166;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14

|||||

Db 52 TTGCCCCGCCCTT 65

RESULT 12

AA430712

LOCUS

DEFINITION zw52g01.r1 Soares total fetus Nb2HF8.9w Homo sapiens cDNA clone IMAGE:773712 5' similar to TR:G309181 G309181 PROCOLLAGEN TYPE V ALPHA 2.; mRNA sequence.

AA430712 1 GI:2111972

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterson, R. and Wilson, R.

WashU-Merck EST Project 1997

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -28m3 rev2 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1..175

/organism="Homo sapiens"

/mol_type="mRNA"

FEATURES

source

ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.1e+04;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/db_xref="taxon:9606"

/clone="IMAGE:773712"

/dev_stage="8-9 weeks"

/lab_host="DH10B"

/clone_lib="Soares total fetus Nb2HF8.9w"

/note="Vector: pT7T3D-PaC (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 14; DB 1; Length 175;

Best Local Similarity 100.0%; Pred. No. 1.1e+04;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14

|||||

Db 2 TTGCCCCGCCCTT 15

RESULT 13

CB721279/c

LOCUS

DEFINITION jnn603501.f op Ornithodoros porcinus porcinus cDNA 5', mRNA

sequence.

CB721279

CB721279.1 GI:29778421

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Ornithodoros porcinus porcinus

Ornithodoros porcinus porcinus

Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

Parasitiformes; Ixodida; Ixodoidea; Argasidae; Ornithodoros.

1 (bases 1 to 191)

Neilan, J.G., Kutish, G.F., Lu, Z., Burrage, T.G., Zsak, A. and

Rock, D.L.

Sequence analysis of Ornithodoros porcinus whole tick cDNA

libraries

Unpublished (2003)

Contact: Neilan JG

Plum Island Animal Disease Center

US Department of Agriculture, Agricultural Research Service

PO Box 848, Greenport, NY 11944-848, USA

Tel: 631 323 3133

Fax: 631 323 3044

Email: jneilan@piadc.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim.alt option. Vector identified by

cross_match v0.990329 and Lucy v1.17p.

Seq primer: M13 Forward.

Location/Qualifiers

1..191

/organism="Ornithodoros porcinus porcinus"

/mol_type="mRNA"

/sub_species="porcinus"

/db_xref="taxon:34596"

/tissue_type="Whole Tick"

/lab_host="DH10B"

/clone_lib="op"

/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;

Library made from polyA selected RNA, isolated from whole

stage N2 ticks two weeks after swine blood meal."

```

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      58 TTGCCCCGCCCTT 45

RESULT 14
CE643460/c
LOCUS   CE643460.1 linear GSS 29-SEP-2003
DEFINITION tigr-gss-dog-l7000366881551 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION CE643460
VERSION   CE643460.1 GI:36961706
KEYWORDS  GSS.
SOURCE    Canis familiaris (dog)
ORGANISM  Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 198)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
          The dog genome: survey sequencing and comparative analysis
          Science 301 (5641), 1898-1903 (2003)
          22875432
          14512627
COMMENT   Contact: Kirkness EF
          The Institute for Genomic Research
          Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
          Rockville, MD 20850, USA
          Tel: 301-838-0200
          Fax: 301-838-0208
          Email: ekirknes@tigr.org
          Class: shotgun
          Location/Qualifiers
            1..198
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /note="Site 1: BstXI; Libraries were prepared from
            peripheral blood"

ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      37 TTGCCCCGCCCTT 24

RESULT 15
AQ843933/c
LOCUS   AQ843933 linear GSS 25-MAY-2001
DEFINITION LMAJFV1_lm01f12.x4 Leishmania major FV1 random genomic library
Leishmania major genomic clone LMAJFV1_lm01f12 3', genomic survey
sequence.
ACCESSION AQ843933
VERSION   AQ843933.1 GI:6011807
KEYWORDS  GSS.
SOURCE    Leishmania major
          Leishmania major
          Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
          Leishmania.
REFERENCE 1 (bases 1 to 200)
AUTHORS   Akopyants N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
          Kissinger,J.C., Roos,D.S. and Beverley,S.M.
          A survey of the Leishmania major Friedlin strain V1 genome by
          shotgun sequencing: a resource for DNA microarrays and expression
          profiling
          Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
          JOURNAL

MEDLINE 21192569
PUBMED  11295190
COMMENT  Contact: Akopyants, NS / Beverley, SM
          WashU Leishmania Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          Library construction: Natalia S. Akopyants, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          IF using this information please cite:
          N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
          Friedlin strain V1 genome by shotgun sequencing' and the Washington
          University Genome Sequencing Center For information on obtaining
          clone material please contact: Natalia S. Akopyants Ph.D.
          (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
          (beverley@borcim.wustl.edu)
          Seq primer: -40UP from Gibco
          Class: shotgun
          High quality sequence stop: 192.
          Location/Qualifiers
            1..200
            /organism="Leishmania major"
            /mol_type="genomic DNA"
            /strain="Friedlin strain V1"
            /db_xref="taxon:5664"
            /clone="LMAJFV1_lm01f12"
            /lab_host="TOP10 (Invitrogen)"
            /clone_lib="Leishmania major FV1 random genomic library"
            /note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
            Genomic DNA was isolated from stationary phase cells. For
            this library, DNA was sheared to give a tight size
            distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
            polymerase, dephosphorylated with Shrimp Alkaline
            Phosphatase and ligated into pZero-2 vector's EcoRV site."

ORIGIN
Query Match 100.0%; Score 14; DB 8; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      81 TTGCCCCGCCCTT 68

RESULT 16
BU755470/c
LOCUS   BU755470 linear mRNA EST 10-OCT-2002
DEFINITION UI-1-BY0-aia-c-04-0-UI-s1 NCI CGAP Tr1 Mus musculus CDNA clone
UI-1-BY0-aia-c-04-0-UI 3', mRNA sequence.
ACCESSION BU755470
VERSION   BU755470.1 GI:23715172
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 223)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: Dr. Janet Rossent
          cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
          cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Clone distribution information can be obtained
          from Dr. M. Bento Soares, bentso-soares@uiowa.edu
          Seq primer: M13 FORWARD

```



```

FEATURES
source
  POLYA=Yes.
  Location/Qualifiers
    1..223
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UI-1-BY0-ala-c-04-0-UI"
      /tissue_type="trophoblast"
      /lab_host="DH10B (Life Technologies)"
      /clone_lib="NCI CGAP Tr1"
      /note="Organ: Trophoblast; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
NCI_CGAP_Tri is a cDNA library containing the following
tissue(s): Placenta trophoblast. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is GGGTG.
TAG_TISSUE=mouse trophoblast
TAG_LIB=UI-1-BY0
TAG_SEQ=GGTTG"

ORIGIN
Query Match 100.0%; Score 14; DB 5; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
Db 84 TTGCCCCGCCCTT 71

RESULT 17
LOCUS CN114354/c 224 bp mRNA linear EST 31-MAR-2004
DEFINITION EC2CAA40DG10.g1 Xenopus tropicalis xthr plasmid library Xenopus
tropicalis cDNA clone xthr40N20 5', mRNA sequence.
ACCESSION CN114354
VERSION CN114354.1 GI:45907050
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
  Xenopodinae; Xenopus; Silurana.
REFERENCE
  1 (bases 1 to 224)
  Thuret R., Fierro A.C., Coen L., Perron M., Demeneix B., Wegnez M.,
  Gyapay G., Weissenbach J., Wincker P., Mazabraud A. and Pollet N.
  Exploring the nervous system transcriptome in the model Xenopus
  tropicalis using EST analysis
  Unpublished (2004)
  Contact: Pollet N
  Transgenese et Genetique des Amphibiens
  CNRS UMR 8080
  IBAIC bat 447, Universite Paris Sud, Orsay, F-91405, France
  Tel: +33 169157272
  Fax: +33 169156816
  Email: Nicolas.Pollet@ibaic.u-psud.fr.

FEATURES
source
  1..224
    /organism="Xenopus tropicalis"
    /mol_type="mRNA"
    /strain="ivory coast"
    /db_xref="taxon:8364"
    /clone="xthr40N20"
    /tissue_type="pool of heads and retinas from tailbud
    stages 25-35"

FEATURES
source
  1..223
    /dev_stage="stage 25-35"
    /lab_host="E.coli DH10B"
    /clone_lib="Xenopus tropicalis xthr plasmid library"
    /note="Vector: pCMVSPORT6 xthr; Site_1: SfiI; Site_2:
SfiI; Xenopus tropicalis polyA+ RNA was obtained from pool
of heads and retinas from tailbud stages 25-35 cDNAs were
synthesized using the SMART system of CLONTECH and
directionally cloned into pCMVSPORT6_xthr, a modified
version of pCMVSPORT6 allowing directional cloning using
asymmetric SfiI sites. For antisense RNA synthesis, use T7
promoter and for sense RNA use Sp6 promoter. Library
constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A.
Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay,
France)."
```

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 224;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 87 TTGCCCCGCCCTT 74

RESULT 18

CNS08WOU/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..240

/organism="Anopheles gambiae"

/mol_type="mRNA"

/strain="6-9"

/db_xref="taxon:7165"

/clone="FK0AAA47BE03"

/plasmid="pME18S-PL"

/note="end : 5-PRIME"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 19

LOCUS

DEFINITION

ACCESSION

VERSION

FEATURES

source

1..224

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="ivory coast"

/db_xref="taxon:8364"

/clone="xthr40N20"

/tissue_type="pool of heads and retinas from tailbud

stages 25-35"

ORIGIN

Query Match

Best Local Similarity

Matches

QY

JOURNAL
COMMENT

Unpublished (2000)
Contact: GSSs: L448k.d.HygT7a.1
Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@bri.org
Seq primer: HygT3
Class: cosmid ends.

FEATURES
source

Location/Qualifiers
1..250
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L448k"
/lab_host="E. coli ED8767"
/clone_lib="Leishmania major Friedlin Cosmid Genomic Library"

/note="Vector: cLHYG; Site 1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sau3AI, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivens et al., Genomics Research, 8:135-145 (1998). The cLHYG vector (Acc. No. CVU59231) is described in Ryan et al., Gene, 131:145-150 (1993)"

ORIGIN

Query Match 100.0%; Score 14; DB 8; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 150 TTGCCCCGCCCTT 163

RESULT 22

A1909047/c

LOCUS A1909047 255 bp mRNA linear EST 30-MAR-2000
DEFINITION QV-BT197-070599-142 BT197 Homo sapiens cDNA, mRNA sequence.
ACCESSION A1909047
VERSION A1909047.1 GI:64999727
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED

20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?tl=QV&t2=QV-BT197-142.html)

FEATURES
source

Seq primer: puc 18 forward.
Location/Qualifiers
1..255
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="female"
/dev_stage="Adult"
/clone_lib="BT197"

/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 14; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 203 TTGCCCCGCCCTT 190

RESULT 23

AA200152

LOCUS AA200152 265 bp mRNA linear EST 19-FEB-1997
DEFINITION mu10b05.r1 Soares thymus_2NBMt Mus musculus cDNA clone IMAGE:638961
5' similar to gb:D63707 Mouse mRNA for hepatoma derived growth factor (MOUSE);, mRNA sequence.

ACCESSION AA200152
VERSION AA200152.1 GI:1796353
KEYWORDS EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 265)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

COMMENT

Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:390953

Seq primer: -28M13 rev2 from Anersham

High quality sequence stop: 248.

FEATURES

source

Location/Qualifiers
1..265
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:638961"
/sex="male"
/tissue type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_thymus_2NBMt"

/note="Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCATCTGAAGTGGAGCGCGCTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 14; DB 1; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCGCGCCCTT 14
|||
Db 184 TTGCCCGCGCCCTT 197

RESULT 24

BF710971/c
LOCUS BF710971 267 bp mRNA linear EST 02-JAN-2001
DEFINITION MI-P-AV1-nrb-c-08-0-UI-s1 MI-P-AV1 Sus scrofa cDNA clone
MI-P-AV1-nrb-c-08-0-UI 3', mRNA sequence.

ACCESSION BF710971
VERSION BF710971.1 GI:12010448
KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 267)

REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Tuggle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: ckuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized placenta library cDNA Library Preparation: M.B. Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 243-265,

>GC_rich#low complexity

Seq primer: MI3 Forward

POLYA=Yes.

FEATURES

source

1..267
Location/Qualifiers
/organism="Sus scrofa"
/mol_type="mRNA"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-AV1-nrb-c-08-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="MI-P-AV1"

/note="Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-AV1 library is normalized library derived from the MI-P-AV0 library, ultimately derived from placenta tissue. For a

detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)

TAG TISSUE=placenta
TAG_LIB=MI-P-AV1
TAG_SEQ=ATTGG"

ORIGIN

Query Match 100.0%; Score 14; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCGCGCCCTT 14
|||
Db 200 TTGCCCGCGCCCTT 187

RESULT 25

BI451845/c

LOCUS BI451845 267 bp mRNA linear EST 21-AUG-2001
DEFINITION TY18E04 hepatocellular carcinoma expression library Marmota monax
cDNA clone TY18E04, mRNA sequence.

ACCESSION BI451845

VERSION BI451845.1 GI:15276552

KEYWORDS EST.

SOURCE Marmota monax (woodchuck)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
Marmota.

1 (bases 1 to 267)

REFERENCE Yamamoto,T., Hazel,C., Yerges,L. and Mason,W.S.

Hepatocellular carcinoma/liver differential expression libraries

Unpublished (2001)

Contact: Mason WS

Fox Chase Cancer Center

7701 Burholme Avenue, Philadelphia, PA 19111, USA

Tel: 215 728 2462

Fax: 215 728 3105

Email: ws.mason@fccc.edu

Similar to Homo sapiens isocitrate dehydrogenase 3 (NAD+) beta (IDH3B), mRNA, accession number XM_044920, with a 255 base overlap(s) with a homology of 94 percent.

Insert Length: 267 Std Error: 0.00

High quality sequence stop: 267.

Location/Qualifiers

source

1..267
/organism="Marmota monax"
/mol_type="mRNA"
/db_xref="taxon:9995"
/clone="TY18E04"

/tissue type="Hepatocellular carcinoma"
/clone_lib="hepatocellular carcinoma expression library"
/note="Organ: Liver: cDNAs were synthesized by reverse transcription of mRNAs purified from the liver and a hepatocellular carcinoma (HCC) of a woodchuck (marmota monax) chronically infected with woodchuck hepatitis virus. cDNAs were digested with RsaI and ligated to one of two different adaptors. HCC enriched cDNA was then generated using the Clontech PCR Select cDNA subtraction kit. Amplicons with non-identical adaptors at each end were selectively amplified by PCR, using primers specific to the adaptors, and the resulting amplicons were ligated into pGEM T easy (Promega)."

ORIGIN

Query Match 100.0%; Score 14; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 TTGCCCCGCCCTT 14
Db      184 TTGCCCCGCCCTT 171

RESULT 26
CF094505      283 bp mRNA linear EST 22-JUL-2003
LOCUS      QHN18C01.YG.abl QH N sunflower H. argophyllus (drought stress)
DEFINITION      Helianthus argophyllus cDNA clone QHN18C01, mRNA sequence.
ACCESSION      CF094505
VERSION
KEYWORDS
SOURCE      EST:
            CF094505.1 GI:33133572
ORGANISM      Helianthus argophyllus
            Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE      1 (bases 1 to 283)
AUTHORS      Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
            Lin, H., van Damme, M., Lavelle, D., Chevaller, P., Ziegler, J.,
            Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
            Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE      Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
JOURNAL      Unpublished (2002)
COMMENT      Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@catc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig1015, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHN18 row: C column: 01.

FEATURES
    source
    1..283
    /organism="Helianthus argophyllus"
    /mol_type="mRNA"
    /db_xref="taxon:73275"
    /lab_host="E.coli"
    /clone="QH18C01"
    /clone_lib="QH N sunflower H. argophyllus (drought stress)"
    /note="Vector: pGEM-T. The library was constructed from
    three different sources (seedling, root and leaf) of RNA
    from a single genotype. cDNAs were pooled and cloned into
    a high-copy vector pGEM-T. Details of library construction
    can be obtained at http://cgpdb.ucdavis.edu/"

ORIGIN
    Query Match      100.0%; Score 14; DB 7; Length 283;
    Best Local Similarity 100.0%; Pred.No. 1.1e+04;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
Db      107 TTGCCCCGCCCTT 120

RESULT 28
CV150513
LOCUS      Mdst6009k09.y1 Mdst Malus x domestica cDNA clone Mdst6009k09 5',
DEFINITION      mRNA sequence.
ACCESSION      CV150513
VERSION      CV150513.1 GI:51912834
KEYWORDS
SOURCE      EST:
            Malus x domestica (cultivated apple)
ORGANISM      Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurousids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE      1 (bases 1 to 286)
AUTHORS      Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,
            Aldwinckle, H., Malnoy, M., Carroll, N., Goldsbrough, P., Orvis, K.,
            Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,
            Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,
            Tsagarisvili, R., Kennedy, S., Waterston, R., and Wilson, R.
            Apple Functional Genomics grant - NSF 0321702
            Unpublished (2004)
            Contact: Schuyler S. Korban
            Apple Functional Genomics grant - NSF 0321702
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library materials provided by: Schuyler S. Korban Library
            constructed by: K. Gasic Library sequenced by: Washington
            University Genome Sequencing Center
            WashU EST name: aao38f05.y1
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.

FEATURES
    source
    1..286
    /organism="Malus x domestica"
    /mol_type="mRNA"

```

```

rosids; eurousids I; Rosales; Rosaceae; Maloideae; Malus.
1 (bases 1 to 286)
Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B.,
McArtney, S., Newcomb, R., Ross, G., Snowden, K., Walton, E. and Yauk, Y.
HortResearch Apple EST Project
Unpublished (2004)
Contact: Gleave, A.
Sequencing Facility
The Horticulture and Food Research Institute of New Zealand Ltd
120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
Tel: 00 64 09 815 4200
Fax: 00 64 09 815 4201
Email: est@hortresearch.co.nz.
Location/Qualifiers
1..286
/organism="Malus x domestica"
/mol_type="mRNA"
/db_xref="taxon:3750"
/clone="AALA003597"
/tissue_type="Fruit cortex"
/dev_stages="150 days after full bloom"
/clone_lib="(AALA) Royal Gala 150 DAFB fruit cortex"
/note="Vector: pBluescript SK(-); Library sequenced by
Genesis Research & Development"

ORIGIN
    Query Match      100.0%; Score 14; DB 7; Length 286;
    Best Local Similarity 100.0%; Pred.No. 1.1e+04;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
Db      107 TTGCCCCGCCCTT 120

RESULT 28
CV150513
LOCUS      Mdst6009k09.y1 Mdst Malus x domestica cDNA clone Mdst6009k09 5',
DEFINITION      mRNA sequence.
ACCESSION      CV150513
VERSION      CV150513.1 GI:51912834
KEYWORDS
SOURCE      EST:
            Malus x domestica (cultivated apple)
ORGANISM      Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurousids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE      1 (bases 1 to 286)
AUTHORS      Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,
            Aldwinckle, H., Malnoy, M., Carroll, N., Goldsbrough, P., Orvis, K.,
            Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,
            Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,
            Tsagarisvili, R., Kennedy, S., Waterston, R., and Wilson, R.
            Apple Functional Genomics grant - NSF 0321702
            Unpublished (2004)
            Contact: Schuyler S. Korban
            Apple Functional Genomics grant - NSF 0321702
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library materials provided by: Schuyler S. Korban Library
            constructed by: K. Gasic Library sequenced by: Washington
            University Genome Sequencing Center
            WashU EST name: aao38f05.y1
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.

FEATURES
    source
    1..286
    /organism="Malus x domestica"
    /mol_type="mRNA"

```

/cultivar="GoldRush"
 /db_xref="taxon:3750"
 /clone="Wdmt6009k09"
 /lab_host="PH10B ampicillin resistant"
 /clone_lib="Wdmt"
 /note="Vector: pBluescript II SK (+); Site_1: NotI;
 Site_2: EcoRII; Total RNA was extracted separately from
 each stage [dormant shoot internodes, active shoot
 internodes, actively-growing shoot internodes], using the
 'pine tree' method. Poly(A)⁺mRNA was isolated twice from
 total RNA from each stage using the Oligotex Direct mRNA
 kit (Qiagen). mRNA was reverse transcribed into double
 stranded cDNA using a modified oligo(dT) primer with an
 identifying tag sequence (see table below). cDNA's from
 different stages were pooled in equal amounts before
 adaptor ligation. Tag identification when sequencing from
 5' end: Stage 1 (dormant shoot internodes) insert
 18(A)TCGTG; Stage 2 (active shoot internodes) insert
 18(A)TGGTG; Stage 3 (actively-growing shoot internodes)
 insert 18(A)TCGTG; Tag identification when sequencing from
 3' end: Stage 1 (dormant shoot internodes) CAGCA18(T)
 insert; Stage 2 (active shoot internodes) CAGCA18(T)
 insert; Stage 3 (actively-growing shoot internodes)
 ACCGA18(T) insert; Double stranded cDNAs were size
 selected (more than 450 bp), adaptored with EcoRI adapters
 at both ends and then digested with NotI. The cDNAs were
 then directionally cloned into EcoRI-NotI digested pBS II
 SK(+) phagemid vector(Stratagene). Identification of
 adaptors and tags in 5'-end sequenced clones:
 <vector>...TAAAGCTT<End Vector><Start
 EcoRI adaptor>GATTCGAATTCATGTGTGGG <End
 Insert> <Start Tag>TCGA<End Tag><Start
 NotI site/Vector>GCGCGCCACGCGGG... The total number of
 white colony forming units (cfu) in the primary library
 before amplification was 2.7x10⁷ cfu (colony forming
 units). The background of empty clones was less than 2%.
 Inserts ranged from 0.5kb to 4 kb, as determined by PCR.
 Purified plasmid DNA from the primary library was
 converted to single-stranded circles and used as a
 template for PCR amplification using the T7 and T3 priming
 sites flanking the cloned cDNA inserts. The purified PCR
 products, representing the entire cloned cDNA population,
 were used as a driver for normalization. Hybridization
 between the single-stranded library and the PCR products
 was carried out for 44 hours at 30C. Unhybridized
 single-stranded DNA circles were separated from hybridized
 DNA rendered partially double-stranded and electroporated
 into DH10B cells to generate the normalized library. The
 total number of clones with insert was 8x10⁵ cfu.
 Background of empty clones was less than 1%."

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 198 TTGCCCCGCCCTT 211

RESULT 29
 CC807244/c 288 bp DNA linear GSS 14-JUL-2003
 DEFINITION ZMMBBc048K19f ZMMBBc Zea mays genomic clone ZMMBBc048K19 5',
 genomic survey sequence.
 ACCESSION CC807244
 VERSION CC807244.1 GI:32637485
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 288)
 AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
 Rouzard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
 TITLE Sequencing of the maize genome at PGIR (2003b)
 JOURNAL Unpublished (2003)
 COMMENT Contact: Bharti,A.K.
 Dr. Joachim Messing's lab
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bharti@waksman.rutgers.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 61.

FEATURES
 source
 1..288
 Location/Qualifiers
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBBc048K19"
 /lab_host="E. coli DH10B"
 /clone_lib="ZMMBBc"
 /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"

ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 288;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 72 TTGCCCCGCCCTT 59

RESULT 30

BG358747 292 bp mRNA linear EST 10-MAR-2001
 LOCUS BOVMS1-002-Q1-E1-A3 Monsanto bovine skeletal muscle cDNA library
 DEFINITION BOVMS1 Bos taurus cDNA clone BOVMS1-002-Q1-E1-A3 5', mRNA sequence.
 ACCESSION BG358747
 VERSION BG358747.1 GI:13271592
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.

REFERENCE

1 (bases 1 to 292)
 AUTHORS Warren,W.C. and Tao,N.
 TITLE Survey of bovine skeletal muscle EST sequences
 JOURNAL Unpublished (2001)
 COMMENT Contact: Wesley C. Warren
 Animal Ag Genomics Group
 Monsanto
 700 Chesterfield Pkwy N., Chesterfield, MO 63198, USA
 Tel: 6367376733
 Fax: 6367377002
 Email: wesley.c.warren@monsanto.com
 Insert Length: 292 Std Error: 0.00
 Seq primer: T7.
 Location/Qualifiers
 1..292
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="BOVMS1-002-Q1-E1-A3"
 /sex="female"
 /tissue_type="skeletal muscle"

FEATURES

source
 1..292
 Location/Qualifiers
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="BOVMS1-002-Q1-E1-A3"
 /sex="female"
 /tissue_type="skeletal muscle"

```

/dev_stage="adult"
/clone_lib="Monsanto bovine skeletal muscle cDNA library
BOVMS1"
/notes="Vector: pSPORT1; Monsanto bovine skeletal muscle
EST sequences"

ORIGIN
Query Match      100.0%; Score 14; DB 4; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
DB 51 TTGCCCCGCCCTT 64

RESULT 31
CR383121
DEFINITION      CR383121 296 bp mRNA linear EST 04-MAY-2004
clone bcas0004a.c.01 5prim, mRNA sequence.
ACCESSION      CR383121
VERSION
KEYWORDS
SOURCE
ORGANISM      Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 296)
Moore,S.S., Li,C., Fu,A., Kneeland,J., Meng,Y., Murdoch,G.,
Dixon,W. and Christopherson,B.
CDNA's from bovine mixed skeletal muscle tissue
Unpublished (2000)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
The sequence best matches gb:HUMCKMA (Human creatine kinase M mRNA,
complete cds. 11/1994) in GenBank main database at E-value of
1e-20.
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 300
POLYA=Yes.
Location/Qualifiers
1..296
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="two males and one female mixed"
/tissue_type="masseter, longissimuslorsi, biceps femoris,
semimembranes"
/cell_type="muscle"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF-strain"
/clone_lib="Bovine Mixed Skeletal Muscle cDNA Library"
/notes="Organ: Skeletal muscle; Vector: Uni-2ZAPXR; Site_1:
ECOR I; Site_2: Xho I"

FEATURES
source
1..296
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="two males and one female mixed"
/tissue_type="masseter, longissimuslorsi, biceps femoris,
semimembranes"
/cell_type="muscle"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF-strain"
/clone_lib="Bovine Mixed Skeletal Muscle cDNA Library"
/notes="Organ: Skeletal muscle; Vector: Uni-2ZAPXR; Site_1:
ECOR I; Site_2: Xho I"

ORIGIN
Query Match      100.0%; Score 14; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
DB 35 TTGCCCCGCCCTT 48

RESULT 33
W14598
LOCUS
DEFINITION      W14598 300 bp mRNA linear EST 10-SEP-1996
IMAGE:330720 5', mRNA sequence.
ACCESSION      W14598
VERSION
KEYWORDS
SOURCE
ORGANISM      Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 300)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morrie,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,I., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

ORIGIN
Query Match      100.0%; Score 14; DB 7; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
DB 22 TTGCCCCGCCCTT 35

RESULT 32
BF261295
LOCUS
DEFINITION      BF261295 300 bp mRNA linear EST 16-NOV-2000
1M0007H12a Bovine Mixed Skeletal Muscle cDNA Library Bos taurus

```

```

CDNA 5', mRNA sequence.
BF261295
VERSION      BF261295.1 GI:11190406
KEYWORDS
SOURCE
ORGANISM      Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 300)
Moore,S.S., Li,C., Fu,A., Kneeland,J., Meng,Y., Murdoch,G.,
Dixon,W. and Christopherson,B.
CDNA's from bovine mixed skeletal muscle tissue
Unpublished (2000)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
The sequence best matches gb:HUMCKMA (Human creatine kinase M mRNA,
complete cds. 11/1994) in GenBank main database at E-value of
1e-20.
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 300
POLYA=Yes.
Location/Qualifiers
1..300
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="two males and one female mixed"
/tissue_type="masseter, longissimuslorsi, biceps femoris,
semimembranes"
/cell_type="muscle"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF-strain"
/clone_lib="Bovine Mixed Skeletal Muscle cDNA Library"
/notes="Organ: Skeletal muscle; Vector: Uni-2ZAPXR; Site_1:
ECOR I; Site_2: Xho I"

ORIGIN
Query Match      100.0%; Score 14; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
DB 35 TTGCCCCGCCCTT 48

RESULT 33
W14598
LOCUS
DEFINITION      W14598 300 bp mRNA linear EST 10-SEP-1996
IMAGE:330720 5', mRNA sequence.
ACCESSION      W14598
VERSION
KEYWORDS
SOURCE
ORGANISM      Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 300)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morrie,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,I., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

```

TITLE JOURNAL COMMENT

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:212120

Seq primer: mob.REGA+ET
High quality sequence stop: 277.

FEATURES

source

Location/Qualifiers
1..300
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:330720"
/dev_stage="19.5 dpc total fetus"
/lab_host="PH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMF19.5"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
Db 83 TTGCCCCGCCCTT 96

RESULT 34

AA136032/c

LOCUS 314 bp mRNA linear EST 01-AUG-1997
DEFINITION zn8e08.r1 Stratagene lung carcinoma 937218 Homo sapiens cDNA clone
IMAGE:565286 5' similar to TR:E124071 E124071 NAD+-ISOCITRATE
DEHYDROGENASE ;, mRNA sequence.

ACCESSION AA136032.1 GI:1697451

VERSION AA136032.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 314)

AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoso,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schallenberg,K., Soares,M.B., Tan,P., Thierry-Mieg,J.,
Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,K.
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

PUBMED 8889549

COMMENT Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1116 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 218.

FEATURES

source

Location/Qualifiers
1..314
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:4594351"
/db_xref="taxon:9606"
/clone="IMAGE:565286"
/issue_type="lung carcinoma"
/cell_line="NCI-H69"
/dev_stage="cell line NCI-H69"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene lung carcinoma 937218"
/note="Organ: lung; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Small cell carcinoma cell line NCI-H69. Average
insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 100.0%; Score 14; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
|||||
Db 113 TTGCCCCGCCCTT 100

RESULT 35

BU028163/c

LOCUS 314 bp mRNA linear EST 23-AUG-2002

DEFINITION QH10K10.Yg.ab1 QH_EFGHJ sunflower RHA280 Helianthus annuus cDNA
clone QH10K10, mRNA sequence.

ACCESSION BU028163

VERSION BU028163.1 GI:22463683

KEYWORDS EST.

SOURCE Helianthus annuus (common sunflower)

ORGANISM Helianthus annuus

REFERENCE 1 (bases 1 to 314)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.

TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compagnonics.ucdavis.edu/

JOURNAL Unpublished (2002)

COMMENT Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

singleton, see http://cgpdb.ucdavis.edu/ for details.

Plate: QH10 row: K column: 10.

Location/Qualifiers

1..314

/organism="Helianthus annuus"

/mol_type="mRNA"

/cultivar="RHA280"

/db_xref="taxon:4232"
 /clone="QH10K10"
 /lab_host="E.coli"
 /clone_lib="QH_EFGHJ sunflower RHA280"
 /note="Vector: pBRCDNASfiAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/TAG_SEQ=Not found"

ORIGIN

Query Match 100.0%; Score 14; DB 5; Length 314;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 56 TTGCCCCGCCCTT 43

RESULT 36

AA033451

LOCUS

DEFINITION mi42a06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
 clone IMAGE:466162 5', mRNA sequence.

ACCESSION AA033451

VERSION AA033451.1

KEYWORDS GI:1504879

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE 1 (bases 1 to 317)
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, P., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WASHU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLM; contact the IMAGE Consortium (info@image.lln.gov) for further information.
 MGI:279978

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLM; contact the IMAGE Consortium (info@image.lln.gov) for further information.
 MGI:279978

FEATURES

source

1..317
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:466162"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH108"

/clone_lib="Soares mouse embryo NbME13.5 14.5"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site1: Not I; Site2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAATGGGACCGCGCGGAATTTTTTTTTTTTTTTTTT

T 3'), on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "

ORIGIN

Query Match 100.0%; Score 14; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 254 TTGCCCCGCCCTT 267

RESULT 37

CN863792

LOCUS

DEFINITION 000919AALA005691HT (AALA) Royal Gala 150 DAFB fruit cortex Malus x domestica cDNA clone AALA005691, mRNA sequence.

ACCESSION CN863792

VERSION CN863792.1

KEYWORDS GI:48120443

SOURCE EST.

ORGANISM Malus x domestica (cultivated apple)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.

REFERENCE

1 (bases 1 to 327)

AUTHORS

Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B., McAtney, S., Newcomb, R., Ross, G., Snowden, K., Walton, B. and Yauk, Y.
 HortResearch Apple EST Project
 Unpublished (2004)
 CONTACT: Gleave, A.
 Sequencing Facility
 The Horticulture and Food Research Institute of New Zealand Ltd
 120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
 Tel: 00 64 09 815 4200
 Fax: 00 64 09 815 4201
 Email: est@hortresearch.co.nz.

TITLE

The Horticulture and Food Research Institute of New Zealand Ltd

JOURNAL

Unpublished (2004)

COMMENT

Sequencing Facility
 The Horticulture and Food Research Institute of New Zealand Ltd
 120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
 Tel: 00 64 09 815 4200
 Fax: 00 64 09 815 4201
 Email: est@hortresearch.co.nz.

FEATURES

source

1..327
 Location/Qualifiers
 /organism="Malus x domestica"
 /mol_type="mRNA"
 /db_xref="taxon:3750"
 /clone="AALA005691"
 /tissue_type="Fruit cortex"
 /dev_stage="150 days after full bloom"
 /clone_lib="(AALA) Royal Gala 150 DAFB fruit cortex"
 /note="Vector: pBluescript SK(-); Library sequenced by Genesis Research & Development"

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 |||||
 Db 123 TTGCCCCGCCCTT 136

RESULT 38

R00632/c

LOCUS

DEFINITION R00632
 Ye74g06.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:123514 5' similar to SP:S40020 S40020 ISOCITRATE
 DEHYDROGENASE ;, mRNA sequence.

ACCESSION R00632

VERSION R00632.1 GI:750368
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 327)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rong, F., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 748
 High quality sequence stops: 224 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 748 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 224.

FEATURES
 source
 1..327
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:476059"
 /db_xref="taxon:9606"
 /clone="IMAGE:123514"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal liver spleen INFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo."

ORIGIN
 Query Match 100.0%; Score 14; DB 7; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCCGCCCTT 14
 |||||
 Db 70 TTGCCCCCGCCCTT 57

ORIGIN
 Query Match 100.0%; Score 14; DB 5; Length 329;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGCCCCCGCCCTT 14
 |||||
 Db 173 TTGCCCCCGCCCTT 186

RESULT 39
BU215685
LOCUS 329 bp mRNA linear EST 25-NOV-2002
DEFINITION 603106847F1 CSEQCHN04 Gallus gallus cdna clone CHEST47m12 5', mRNA sequence.
ACCESSION BU215685
VERSION BU215685.1 GI:25393974
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 329)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,

TITLE Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
JOURNAL A Comprehensive Collection of Chicken cDNAs
SOURCE Curr. Biol. 12 (22), 1965-1969 (2002)
ORGANISM 2235534
COMMENT 12445392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
FEATURES
 Location/Qualifiers
 1..329
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST47m12"
 /tissue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN04"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research (1996): 791, except that a significantly longer reannealing hybridization was used."

TITLE AA030175
JOURNAL 331 bp mRNA linear EST 21-JAN-1997
COMMENT mh88c01.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus CDNA clone IMAGE:458016 5', mRNA sequence.
ACCESSION AA030175
VERSION AA030175.1 GI:1497313
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 331)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellengberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501 St. Louis, MO 63108

10409428
 PUBMED
 COMMENT
 Contact: Jeffery L. Miller
 Laboratory of Chemical Biology
 National Institute of Diabetes and Digestive and Kidney Diseases
 Building 10, Room 9B17, National Institutes of Health, Bethesda, MD
 20892, USA
 Tel: 301 402 2373
 Fax: 301 435 5148
 Email: jm7@nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 18 row: a column: 05
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers
 1..336
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="ax18a05"
 /sex="unknown"
 /tissue_type="blood"
 /cell_type="Erythroid Cells"
 /cell_line="Primary Culture of Peripheral Blood
 Mononuclear Cells"
 /dev_stages="Progenitor; EPO responsive CD71+++"
 /lab_host="SOLR"
 /clone_lib="Hembase; Erythroid Progenitor Cells (LCB:ax
 library)"
 /note="Organ: blood; Vector: Lambda ZAP II; Site 1: EcoRI;
 Site 2: EcoRI; 65,000 proliferating erythroid cells from
 the buffy coat of a blood donation were obtained by flow
 cytometric separation after a 5-day culture period in the
 presence of erythropoietin. Total RNA was purified from
 the sorted cell population using RNeasy reagent. RNA (0.3
 ug) was converted into double stranded cDNA using
 Clontech's CapFinder cDNA Library Construction Kit
 (Clontech) according to the manufacturer's protocol and
 cloned into EcoRI digested Lambda Zap II vector
 (Stratagene). The phage library was amplified once prior
 to in vivo excision in SOLR cells. Individual colonies
 were grown, and the cDNA inserts were sequenced in high
 throughput (NIH intramural sequencing center
 http://www.nisc.nih.gov/)."

Query Match 100.0%; Score 14; DB 4; Length 336;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
 |||||
 DB 128 TTGCCCCCGCCCTT 115

RESULT 43
 CC852229/c
 LOCUS
 DEFINITION
 CC852229
 NOTRE Dame Liverpool-101M13, genomic survey sequence.
 CC852229
 GSS.
 CC852229.1 GI:33204910
 Aedes aegypti (yellow fever mosquito)
 Aedes aegypti
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
 Stegomyia.
 1 (bases 1 to 338)
 Loftus B., Shetty, J., Knudson, D. and Severson, D.
 BAC end sequencing of Aedes aegypti
 unpublished (2003)
 Other GSSs: NDL:101M13.T7
 Contact: Brendan Loftus

10409428
 PUBMED
 COMMENT
 Contact: Jeffery L. Miller
 Laboratory of Chemical Biology
 National Institute of Diabetes and Digestive and Kidney Diseases
 Building 10, Room 9B17, National Institutes of Health, Bethesda, MD
 20892, USA
 Tel: 301 402 2373
 Fax: 301 435 5148
 Email: jm7@nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 18 row: a column: 05
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers
 1..336
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="ax18a05"
 /sex="unknown"
 /tissue_type="blood"
 /cell_type="Erythroid Cells"
 /cell_line="Primary Culture of Peripheral Blood
 Mononuclear Cells"
 /dev_stages="Progenitor; EPO responsive CD71+++"
 /lab_host="SOLR"
 /clone_lib="Hembase; Erythroid Progenitor Cells (LCB:ax
 library)"
 /note="Organ: blood; Vector: Lambda ZAP II; Site 1: EcoRI;
 Site 2: EcoRI; 65,000 proliferating erythroid cells from
 the buffy coat of a blood donation were obtained by flow
 cytometric separation after a 5-day culture period in the
 presence of erythropoietin. Total RNA was purified from
 the sorted cell population using RNeasy reagent. RNA (0.3
 ug) was converted into double stranded cDNA using
 Clontech's CapFinder cDNA Library Construction Kit
 (Clontech) according to the manufacturer's protocol and
 cloned into EcoRI digested Lambda Zap II vector
 (Stratagene). The phage library was amplified once prior
 to in vivo excision in SOLR cells. Individual colonies
 were grown, and the cDNA inserts were sequenced in high
 throughput (NIH intramural sequencing center
 http://www.nisc.nih.gov/)."

Query Match 100.0%; Score 14; DB 4; Length 336;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
 |||||
 DB 128 TTGCCCCCGCCCTT 115

RESULT 43
 CC852229/c
 LOCUS
 DEFINITION
 CC852229
 NOTRE Dame Liverpool-101M13, genomic survey sequence.
 CC852229
 GSS.
 CC852229.1 GI:33204910
 Aedes aegypti (yellow fever mosquito)
 Aedes aegypti
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
 Stegomyia.
 1 (bases 1 to 338)
 Loftus B., Shetty, J., Knudson, D. and Severson, D.
 BAC end sequencing of Aedes aegypti
 unpublished (2003)
 Other GSSs: NDL:101M13.T7
 Contact: Brendan Loftus

Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: enta@tigr.org
 Library was provided by David Severson
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..338
 /organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /strain="Liverpool"
 /db_xref="taxon:7159"
 /clone="Notre Dame Liverpool-101M13"
 /clone_lib="Notre Dame Liverpool"
 /note="Vector: pGCBAC1; Site 1: Hind III; The library was
 prepared from whole body tissue of newly hatched L1 larvae
 by David Severson at the University of Notre Dame and
 Hongbin Zhang"

FEATURES
 source
 1..338
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stages="Adult"
 /clone_lib="KT0031"
 /note="Organ: bladder_tumor; Vector: puc18; Site 1: SmaI;

ORIGIN
 Query Match 100.0%; Score 14; DB 9; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1.1e+04;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
 |||||
 DB 34 TTGCCCCCGCCCTT 21

RESULT 44
 CV318910/c
 LOCUS
 DEFINITION
 CV318910
 CV318910.1 GI:52642124
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 345)
 Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. http://www.ludwig.org.br.
 Location/Qualifiers
 1..345
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stages="Adult"
 /clone_lib="KT0031"

FEATURES
 source
 1..345
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stages="Adult"
 /clone_lib="KT0031"

Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 133 TTGCCCCGCCCTT 120

RESULT 45

AQ106680/c

LOCUS HS_3080_B2_G05_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3080 Col=10 Row=N, genomic survey sequence.
ACCESSION AQ106680.1 GI:3482036
VERSION GSS.
KEYWORDS Homo sapiens (human)
SOURCE GSS.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 345)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

PUBMED 10449764

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3080 row: N column: 10
Class: BAC ends
High quality sequence stop: 345.
Location/Qualifiers

1. 345
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3080 Col=10 Row=N"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

Query Match 100.0%; Score 14; DB 8; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 150 TTGCCCCGCCCTT 137

RESULT 46

BG939965/c

LOCUS BG939965 350 bp mRNA linear EST 20-MAY-2002
DEFINITION Homo sapiens CDNA clone ax01g08 random, mRNA sequence.

ACCESSION BG939965
VERSION BG939965.1 GI:14339348
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 350)
AUTHORS Gubin,A.N., Njoroge,J.M., Bouffard,G.G. and Miller,J.L.
TITLE Gene expression in proliferating human erythroid cells
JOURNAL Genomics 59 (2), 168-177 (1999)
MEDLINE 99339981
PUBMED 10409428
COMMENT Contact: Jeffery L. Miller
Laboratory of Chemical Biology
National Institute of Diabetes and Digestive and Kidney Diseases
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD 20892, USA
Tel: 301 402 2373
Fax: 301 435 5148
Email: jm7f@nih.gov
DNA Sequencing and analyses by National Institutes of Health Intramural Sequencing Center (NISC).
Plate: 01 row: g column: 08
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers

1. 350
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ax01g08"
/sex="unknown"
/tissue_type="blood"
/cell_type="Erythroid Cells"
/cell_lines="Primary Culture of Peripheral Blood Mononuclear Cells"
/dev_stages="Progenitor; EPO responsive CD71++++"
/lab_host="SOLR"
/clone_lib="Hembase; Erythroid Progenitor Cells (LCB:ax library)"
/note="Organ: blood; Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: EcoRI; 65,000 proliferating erythroid cells from the Buffy coat of a blood donation were obtained by flow cytometric separation after a 5-day culture period in the presence of erythropoietin. Total RNA was purified from the sorted cell population using Trizol reagent. RNA (0.3 ug) was converted into double stranded cDNA using Clontech's CapFinder cDNA Library Construction Kit (Clontech) according to the manufacturer's protocol and cloned into EcoRI digested Lambda Zap II vector (Stratagene). The phage library was amplified once prior to in vivo excision in SOLR cells. Individual colonies were grown, and the cDNA inserts were sequenced in high throughput (NIH intramural sequencing center http://www.nisc.nih.gov/)."

Query Match 100.0%; Score 14; DB 4; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 234 TTGCCCCGCCCTT 221

RESULT 47

BQ343488/c

LOCUS BQ343488 350 bp mRNA linear EST 20-MAY-2002
DEFINITION IL3-NT0104-240500-146-B05 NTO104 Homo sapiens cDNA, mRNA sequence.

Query Match 100.0%; Score 14; DB 4; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 234 TTGCCCCGCCCTT 221

```

ACCESSION BQ343488
VERSION BQ343488.1 GI:21007549
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 350)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-NT0104-
240500-146-B05&t3=2000-05-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 350.
FEATURES
source
1..350
Location/Qualifiers
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0104"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196.716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 100.0%; Score 14; DB 5; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
| | | | |
Db 107 TTGCCCCGCCCTT 94
| | | | |
RESULT 48
BG277108/c 351 bp mRNA linear EST 21-FEB-2001
LOCUS IMAGE:3514863 5', mRNA sequence.
DEFINITION BG277108
ACCESSION BG277108.1 GI:13072080
VERSION BG277108
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 351)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: ux63e08.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1395439
Seq primer: -40RP from Gibco
High quality sequence stop: 311.
FEATURES
source
1..351
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:3514863"
/tissue_type="mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares NKWMD mandible"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15',
TGTTCACAACTGTAAGTGGAGCGCCGCTTAATTTTTTTTTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Facima Bonaldo. "
ORIGIN
Query Match 100.0%; Score 14; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
| | | | |
Db 36 TTGCCCCGCCCTT 23
| | | | |
RESULT 49
CR382632 352 bp mRNA linear EST 04-MAY-2004
LOCUS CR382632 Bovine multi-stage muscles library (bcaj) Bos taurus cDNA
DEFINITION clone bcaj0004a.d.03 Sprim, mRNA sequence.
ACCESSION CR382632
VERSION CR382632.1 GI:47004536
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 352)
AUTHORS Chevalet,C.
TITLE AGENAE, a French Animal Genome project
JOURNAL Unpublished (2004)
COMMENT Contact: Martin P
INRA, Genomique & Physiologie de la Lactation
Domaine de Vilvert, 78352 JOUY-EN-JOSAS cedex, FRANCE
Tel: +33 (0) 1.34.65.25.82
Fax: +33 (0) 1.34.65.29.26
Email: Patrice.Martin@jouy.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at signenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
FEATURES
source
Plate: 0004 row: d column: 3.
Location/Qualifiers
1..352
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="bcaj0004a.d.03"

```

Search completed: April 25, 2005, 15:04:50
Job time : 1231.72 secs

/tissue type="muscles : heart, longissimus thoracis,
semitendinosus, masseter, cutaneus trunci"
/dev stage="from embryos to adults"
/clone lib="Bovine multi-stage muscles library (bcaj)"
/note="Clone distribution : AGENAE Resource centre.
Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA
Radiobiologie et Etude du genome (LREG), Domaine de
Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0)
1.34.65.28.02, +33 (0) 1.34.65.22.73"

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 78 TTGCCCCGCCCTT 91

RESULT 50

CR382747
LOCUS 353 bp mRNA linear EST 04-MAY-2004
DEFINITION CR382747 Bovine multi-stage muscles library (bcaj) Bos taurus cDNA
clone bcaj0004a.g.01 5prim, mRNA sequence.
CR382747
ACCESSION CR382747.1 GI:47004651
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 353)
Chevalet, C.
AGENAE, a French Animal Genome project
Unpublished (2004)
Contact: Martin P
INRA, Genomique & Physiologie de la Lactation
Domaine de Vilvert, 78352 JOUY-EN-JOSAS cedex, FRANCE
Tel: +33 (0) 1.34.65.25.82
Fax: +33 (0) 1.34.65.29.26
Email: Patrice.Martin@jouy.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at signenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.

Plate: 0004 row: 9 column: 1.

FEATURES

source
Location/Qualifiers
1..353
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="bcaj0004a.g.01"
/tissue type="muscles : heart, longissimus thoracis,
semitendinosus, masseter, cutaneus trunci"
/dev stages="from embryos to adults"
/clone lib="Bovine multi-stage muscles library (bcaj)"
/note="Clone distribution : AGENAE Resource centre.
Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA
Radiobiologie et Etude du genome (LREG), Domaine de
Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0)
1.34.65.28.02, +33 (0) 1.34.65.22.73"

ORIGIN

Query Match 100.0%; Score 14; DB 7; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
|||||
Db 79 TTGCCCCGCCCTT 92

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:16:14 ; Search time 174.759 Seconds
(without alignments)
486.355 Million cell updates/sec

Title: US-10-010-476-15

Perfect score: 14

Sequence: 1 TTGCCCCGCCCTT 14

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	14	15	US-10-010-476-15
2	14	100.0	22	15	US-10-010-476-14
3	14	100.0	30	15	US-10-010-476-9
4	14	100.0	30	15	US-10-010-476-10
5	14	100.0	50	15	US-10-010-476-13
6	14	100.0	110	17	US-10-242-535A-565
7	14	100.0	110	17	US-10-085-783A-565
8	14	100.0	292	9	US-09-960-352-462
9	14	100.0	326	10	US-09-803-719-2383
10	14	100.0	340	10	US-09-803-719-2387
11	14	100.0	394	18	US-10-425-115-153894
					Sequence 15, Appl
					Sequence 14, Appl
					Sequence 9, Appl
					Sequence 10, Appl
					Sequence 13, Appl
					Sequence 565, Appl
					Sequence 565, Appl
					Sequence 462, Appl
					Sequence 2383, Appl
					Sequence 2387, Appl
					Sequence 153894,

12	14	100.0	526	18	US-10-425-115-131927	Sequence 131927,
13	14	100.0	543	17	US-10-424-599-140509	Sequence 140509,
14	14	100.0	600	18	US-10-437-963-23923	Sequence 23923, A
15	14	100.0	2352	18	US-10-437-963-49030	Sequence 49030, A
16	14	100.0	2573	17	US-10-152-319A-2018	Sequence 2018, Ap
17	14	100.0	2850	17	US-10-133-937-26	Sequence 26, Appl
18	14	100.0	2850	17	US-10-159-563-26	Sequence 26, Appl
19	14	100.0	3170	17	US-10-276-645-1	Sequence 1, Appl
20	14	100.0	3250	17	US-10-276-645-2	Sequence 2, Appl
21	14	100.0	13273	9	US-09-764-869-2349	Sequence 2349, Ap
22	14	100.0	13273	14	US-10-091-504-2349	Sequence 2349, Ap
23	14	100.0	13273	17	US-10-227-577-2349	Sequence 11, Appl
24	14	100.0	51001	17	US-10-189-268-11	Sequence 20955, A
25	13	92.9	201	19	US-10-741-600-20995	Sequence 20995, A
26	13	92.9	201	19	US-10-741-600-21067	Sequence 21067, A
27	13	92.9	201	19	US-10-741-600-21068	Sequence 21068, A
28	13	92.9	207	17	US-10-369-493-35585	Sequence 25585, A
29	13	92.9	260	18	US-10-477-797-2	Sequence 2, Appl
30	13	92.9	260	19	US-10-500-173-2	Sequence 2, Appl
31	13	92.9	265	9	US-09-923-876-326	Sequence 326, App
32	13	92.9	265	10	US-09-923-876-326	Sequence 326, App
33	13	92.9	291	18	US-10-425-115-29238	Sequence 29238, A
34	13	92.9	332	18	US-10-425-115-60113	Sequence 60113, A
35	13	92.9	333	18	US-10-477-797-3	Sequence 3, Appl
36	13	92.9	333	19	US-10-500-173-3	Sequence 3, Appl
37	13	92.9	347	18	US-10-425-115-159870	Sequence 159870,
38	13	92.9	348	18	US-10-425-115-181957	Sequence 181957,
39	13	92.9	392	18	US-10-425-115-177239	Sequence 177239,
40	13	92.9	424	14	US-10-116-802-513	Sequence 513, App
41	13	92.9	467	10	US-09-918-995-21483	Sequence 21483, A
42	13	92.9	475	18	US-10-767-701-5731	Sequence 5731, Ap
43	13	92.9	477	17	US-10-424-599-24118	Sequence 24118, A
44	13	92.9	529	18	US-10-723-860-20	Sequence 20, Appl
45	13	92.9	534	16	US-10-029-386-12456	Sequence 12456, A
46	13	92.9	544	17	US-10-152-319A-867	Sequence 867, App
47	13	92.9	553	18	US-10-425-115-96771	Sequence 96771, A
48	13	92.9	577	18	US-10-767-701-16993	Sequence 16993, A
49	13	92.9	581	9	US-09-864-761-11958	Sequence 11958, A
50	13	92.9	582	18	US-10-425-115-10159	Sequence 10159, A
51	13	92.9	586	13	US-10-027-632-205819	Sequence 205819,
52	13	92.9	586	13	US-10-027-632-205820	Sequence 205820,
53	13	92.9	586	17	US-10-027-632-205819	Sequence 205819,
54	13	92.9	586	17	US-10-027-632-205820	Sequence 205820,
55	13	92.9	604	18	US-10-425-115-58957	Sequence 58957, A
56	13	92.9	617	18	US-10-425-115-14224	Sequence 14224, A
57	13	92.9	621	13	US-10-027-632-270900	Sequence 270900,
58	13	92.9	621	17	US-10-027-632-270900	Sequence 270900,
59	13	92.9	679	11	US-09-889-325-1	Sequence 1, Appl
60	13	92.9	679	18	US-10-882-405-1	Sequence 1, Appl
61	13	92.9	685	13	US-10-027-632-20577	Sequence 20577, A
62	13	92.9	685	17	US-10-027-632-20577	Sequence 20577, A
63	13	92.9	741	15	US-10-259-165-645	Sequence 645, App
64	13	92.9	747	18	US-10-437-963-17037	Sequence 17037, A
65	13	92.9	777	13	US-10-027-632-170022	Sequence 170022,
66	13	92.9	777	17	US-10-027-632-170022	Sequence 170022,
67	13	92.9	793	17	US-10-027-632-163411	Sequence 163411,
68	13	92.9	793	17	US-10-027-632-163411	Sequence 163411,
69	13	92.9	809	13	US-10-027-632-153521	Sequence 153521,
70	13	92.9	809	17	US-10-027-632-153521	Sequence 153521,
71	13	92.9	820	18	US-10-425-115-43277	Sequence 43277, A
72	13	92.9	855	18	US-10-437-963-94331	Sequence 94331, A
73	13	92.9	878	17	US-10-425-115-110827	Sequence 110827,
74	13	92.9	903	17	US-10-260-238-5715	Sequence 5715, Ap
75	13	92.9	969	17	US-10-260-238-5454	Sequence 5454, Ap
76	13	92.9	1015	19	US-10-486-706-107	Sequence 107, App
77	13	92.9	1026	18	US-10-437-963-92392	Sequence 92392, A
78	13	92.9	1065	17	US-10-425-114-23849	Sequence 23849, A
79	13	92.9	1052	17	US-10-369-493-47162	Sequence 47162, A
80	13	92.9	1097	17	US-10-425-114-22462	Sequence 22462, A
81	13	92.9	1099	17	US-10-425-114-1540	Sequence 1540, Ap
82	13	92.9	1104	10	US-09-814-353-21494	Sequence 21494, A
83	13	92.9	1163	18	US-10-437-963-92395	Sequence 92395, A
84	13	92.9	1167	17	US-10-425-114-31222	Sequence 31222, A

85	13	92.9	1185	17	US-10-425-114-15735	Sequence 15735, A	158	13	92.9	1657	10	US-09-992-521-198	Sequence 198, App
86	13	92.9	1195	15	US-10-304-928-7	Sequence 7, Appli	159	13	92.9	1657	10	US-09-997-333-198	Sequence 198, App
87	13	92.9	1199	18	US-10-739-930-2611	Sequence 2611, Ap	160	13	92.9	1657	10	US-09-997-384-198	Sequence 198, App
88	13	92.9	1202	13	US-10-027-632-101097	Sequence 101097,	161	13	92.9	1657	10	US-09-998-041-198	Sequence 198, App
c 88	13	92.9	1202	17	US-10-027-632-101097	Sequence 101097,	162	13	92.9	1657	10	US-09-997-585-198	Sequence 198, App
89	13	92.9	1238	17	US-10-425-114-15826	Sequence 15826, A	163	13	92.9	1657	10	US-09-997-614-198	Sequence 198, App
90	13	92.9	1240	18	US-10-425-115-43278	Sequence 43278, A	164	13	92.9	1657	10	US-09-989-862-198	Sequence 198, App
91	13	92.9	1326	15	US-10-037-270-867	Sequence 867, App	165	13	92.9	1657	10	US-09-997-529-198	Sequence 198, App
92	13	92.9	1326	17	US-10-117-722-867	Sequence 867, App	166	13	92.9	1657	10	US-09-989-725-198	Sequence 198, App
93	13	92.9	1365	17	US-10-425-114-25201	Sequence 25201, A	167	13	92.9	1657	10	US-09-991-550-198	Sequence 198, App
c 93	13	92.9	1404	17	US-10-369-493-46601	Sequence 46601, A	168	13	92.9	1657	10	US-09-997-641-198	Sequence 198, App
c 95	13	92.9	1428	18	US-10-437-963-89437	Sequence 89437, A	169	13	92.9	1657	10	US-09-989-733-198	Sequence 198, App
c 96	13	92.9	1557	18	US-10-425-115-7566	Sequence 7566, Ap	170	13	92.9	1657	10	US-09-992-643-198	Sequence 198, App
97	13	92.9	1603	18	US-10-425-115-119009	Sequence 119009,	171	13	92.9	1657	14	US-10-028-072-433	Sequence 433, App
c 98	13	92.9	1621	18	US-10-767-701-12166	Sequence 12166, A	172	13	92.9	1657	14	US-10-140-808-433	Sequence 433, App
99	13	92.9	1657	9	US-09-989-722-198	Sequence 198, App	173	13	92.9	1657	14	US-10-121-049-433	Sequence 433, App
100	13	92.9	1657	9	US-09-989-723-198	Sequence 198, App	174	13	92.9	1657	14	US-10-123-904-433	Sequence 433, App
101	13	92.9	1657	9	US-09-989-279-198	Sequence 198, App	175	13	92.9	1657	14	US-10-140-470-433	Sequence 433, App
102	13	92.9	1657	9	US-09-989-727-198	Sequence 198, App	176	13	92.9	1657	14	US-10-175-746-433	Sequence 433, App
103	13	92.9	1657	9	US-09-989-731-198	Sequence 198, App	177	13	92.9	1657	14	US-10-176-918-433	Sequence 433, App
104	13	92.9	1657	9	US-09-989-733-198	Sequence 198, App	178	13	92.9	1657	14	US-10-176-921-433	Sequence 433, App
105	13	92.9	1657	9	US-09-991-073-198	Sequence 198, App	179	13	92.9	1657	14	US-10-137-865-433	Sequence 433, App
106	13	92.9	1657	9	US-09-990-442-198	Sequence 198, App	180	13	92.9	1657	14	US-10-140-474-433	Sequence 433, App
107	13	92.9	1657	9	US-09-991-163-198	Sequence 198, App	181	13	92.9	1657	14	US-10-142-431-433	Sequence 433, App
108	13	92.9	1657	9	US-09-993-604-198	Sequence 198, App	182	13	92.9	1657	14	US-10-143-114-433	Sequence 433, App
109	13	92.9	1657	9	US-09-990-456-198	Sequence 198, App	183	13	92.9	1657	14	US-10-142-419-433	Sequence 433, App
110	13	92.9	1657	9	US-09-989-721-198	Sequence 198, App	184	13	92.9	1657	14	US-10-123-262-433	Sequence 433, App
111	13	92.9	1657	9	US-09-992-598-198	Sequence 198, App	185	13	92.9	1657	14	US-10-142-423-433	Sequence 433, App
112	13	92.9	1657	9	US-09-989-293A-198	Sequence 198, App	186	13	92.9	1657	14	US-10-121-050-433	Sequence 433, App
113	13	92.9	1657	9	US-09-989-735-198	Sequence 198, App	187	13	92.9	1657	14	US-10-141-755-433	Sequence 433, App
114	13	92.9	1657	9	US-09-990-444-198	Sequence 198, App	188	13	92.9	1657	14	US-10-143-032-433	Sequence 433, App
115	13	92.9	1657	9	US-09-991-181-198	Sequence 198, App	189	13	92.9	1657	14	US-10-123-108-433	Sequence 433, App
116	13	92.9	1657	9	US-09-989-730-198	Sequence 198, App	190	13	92.9	1657	14	US-10-123-236-433	Sequence 433, App
117	13	92.9	1657	9	US-09-990-436-198	Sequence 198, App	191	13	92.9	1657	14	US-10-123-261-433	Sequence 433, App
118	13	92.9	1657	9	US-09-993-687-198	Sequence 198, App	192	13	92.9	1657	14	US-10-140-921-433	Sequence 433, App
119	13	92.9	1657	10	US-09-989-734-198	Sequence 198, App	193	13	92.9	1657	14	US-10-140-928-433	Sequence 433, App
120	13	92.9	1657	10	US-09-997-653-198	Sequence 198, App	194	13	92.9	1657	14	US-10-121-045-433	Sequence 433, App
121	13	92.9	1657	10	US-09-989-724-198	Sequence 198, App	195	13	92.9	1657	14	US-10-123-292-433	Sequence 433, App
122	13	92.9	1657	10	US-09-989-728-198	Sequence 198, App	196	13	92.9	1657	14	US-10-123-903-433	Sequence 433, App
123	13	92.9	1657	10	US-09-990-441-198	Sequence 198, App	197	13	92.9	1657	14	US-10-124-819-433	Sequence 433, App
124	13	92.9	1657	10	US-09-993-667-198	Sequence 198, App	198	13	92.9	1657	14	US-10-124-822-433	Sequence 433, App
125	13	92.9	1657	10	US-09-997-428-198	Sequence 198, App	199	13	92.9	1657	14	US-10-140-925-433	Sequence 433, App
126	13	92.9	1657	10	US-09-997-666-198	Sequence 198, App	200	13	92.9	1657	14	US-10-160-498-433	Sequence 433, App
127	13	92.9	1657	10	US-09-990-438-198	Sequence 198, App	201	13	92.9	1657	14	US-10-124-824-433	Sequence 433, App
128	13	92.9	1657	10	US-09-990-562-198	Sequence 198, App	202	13	92.9	1657	14	US-10-127-825A-433	Sequence 433, App
129	13	92.9	1657	10	US-09-980-711-198	Sequence 198, App	203	13	92.9	1657	14	US-10-127-829A-433	Sequence 433, App
130	13	92.9	1657	10	US-09-989-726-198	Sequence 198, App	204	13	92.9	1657	14	US-10-127-839A-433	Sequence 433, App
131	13	92.9	1657	10	US-09-990-437-198	Sequence 198, App	205	13	92.9	1657	14	US-10-127-901A-433	Sequence 433, App
132	13	92.9	1657	10	US-09-991-157-198	Sequence 198, App	206	13	92.9	1657	14	US-10-128-693A-433	Sequence 433, App
133	13	92.9	1657	10	US-09-997-514-198	Sequence 198, App	207	13	92.9	1657	14	US-10-131-813A-433	Sequence 433, App
134	13	92.9	1657	10	US-09-997-573-198	Sequence 198, App	208	13	92.9	1657	14	US-10-131-818A-433	Sequence 433, App
135	13	92.9	1657	10	US-09-991-172-198	Sequence 198, App	209	13	92.9	1657	14	US-10-131-823A-433	Sequence 433, App
136	13	92.9	1657	10	US-09-997-628-198	Sequence 198, App	210	13	92.9	1657	14	US-10-131-824A-433	Sequence 433, App
137	13	92.9	1657	10	US-09-997-683-198	Sequence 198, App	211	13	92.9	1657	14	US-10-131-830A-433	Sequence 433, App
138	13	92.9	1657	10	US-09-989-729A-198	Sequence 198, App	212	13	92.9	1657	14	US-10-137-872A-433	Sequence 433, App
139	13	92.9	1657	10	US-09-997-349-198	Sequence 198, App	213	13	92.9	1657	14	US-10-147-502-433	Sequence 433, App
140	13	92.9	1657	10	US-09-997-440-198	Sequence 198, App	214	13	92.9	1657	14	US-10-147-515-433	Sequence 433, App
141	13	92.9	1657	10	US-09-990-443-198	Sequence 198, App	215	13	92.9	1657	14	US-10-147-517-433	Sequence 433, App
142	13	92.9	1657	10	US-09-997-549-198	Sequence 198, App	216	13	92.9	1657	14	US-10-147-527-433	Sequence 433, App
143	13	92.9	1657	10	US-09-990-440-198	Sequence 198, App	217	13	92.9	1657	14	US-10-121-041-433	Sequence 433, App
144	13	92.9	1657	10	US-09-997-857-198	Sequence 198, App	218	13	92.9	1657	14	US-10-121-043-433	Sequence 433, App
145	13	92.9	1657	10	US-09-993-469-198	Sequence 198, App	219	13	92.9	1657	14	US-10-121-047-433	Sequence 433, App
146	13	92.9	1657	10	US-09-997-542-198	Sequence 198, App	220	13	92.9	1657	14	US-10-123-215-433	Sequence 433, App
147	13	92.9	1657	10	US-09-993-748-198	Sequence 198, App	221	13	92.9	1657	14	US-10-123-902-433	Sequence 433, App
148	13	92.9	1657	10	US-09-990-439-198	Sequence 198, App	222	13	92.9	1657	14	US-10-123-908-433	Sequence 433, App
149	13	92.9	1657	10	US-09-990-427-198	Sequence 198, App	223	13	92.9	1657	14	US-10-123-909-433	Sequence 433, App
150	13	92.9	1657	10	US-09-989-328-198	Sequence 198, App	224	13	92.9	1657	14	US-10-124-813-433	Sequence 433, App
151	13	92.9	1657	10	US-09-993-583-198	Sequence 198, App	225	13	92.9	1657	14	US-10-124-817-433	Sequence 433, App
152	13	92.9	1657	10	US-09-991-992-198	Sequence 198, App	226	13	92.9	1657	14		
153	13	92.9	1657	10			227	13	92.9	1657	14		
154	13	92.9	1657	10			228	13	92.9	1657	14		
155	13	92.9	1657	10			229	13	92.9	1657	14		
156	13	92.9	1657	10			230	13	92.9	1657	14		
157	13	92.9	1657	10			231	13	92.9	1657	14		

377	13	92.9	1657	16	US-10-142-424-433	Sequence 433, App	13	92.9	1657	16	US-10-152-403-433	Sequence 433, App
378	13	92.9	1657	16	US-10-142-761-433	Sequence 433, App	13	92.9	1657	16	US-10-153-756-433	Sequence 433, App
379	13	92.9	1657	16	US-10-142-763-433	Sequence 433, App	13	92.9	1657	16	US-10-157-784-433	Sequence 433, App
380	13	92.9	1657	16	US-10-142-765-433	Sequence 433, App	13	92.9	1657	16	US-10-157-797-433	Sequence 433, App
381	13	92.9	1657	16	US-10-142-867-433	Sequence 433, App	13	92.9	1657	16	US-10-158-491-433	Sequence 433, App
382	13	92.9	1657	16	US-10-142-888-433	Sequence 433, App	13	92.9	1657	16	US-10-142-762-433	Sequence 433, App
383	13	92.9	1657	16	US-10-143-034-433	Sequence 433, App	13	92.9	1657	16	US-10-142-764-433	Sequence 433, App
384	13	92.9	1657	16	US-10-143-116-433	Sequence 433, App	13	92.9	1657	16	US-10-142-766-433	Sequence 433, App
385	13	92.9	1657	16	US-10-143-117-433	Sequence 433, App	13	92.9	1657	16	US-10-145-625-433	Sequence 433, App
386	13	92.9	1657	16	US-10-144-957-433	Sequence 433, App	13	92.9	1657	16	US-10-145-627-433	Sequence 433, App
387	13	92.9	1657	16	US-10-144-992-433	Sequence 433, App	13	92.9	1657	16	US-10-145-660-433	Sequence 433, App
388	13	92.9	1657	16	US-10-145-015-433	Sequence 433, App	13	92.9	1657	16	US-10-145-962-433	Sequence 433, App
389	13	92.9	1657	16	US-10-145-090-433	Sequence 433, App	13	92.9	1657	16	US-10-146-789-433	Sequence 433, App
390	13	92.9	1657	16	US-10-145-091-433	Sequence 433, App	13	92.9	1657	16	US-10-147-483-433	Sequence 433, App
391	13	92.9	1657	16	US-10-145-629-433	Sequence 433, App	13	92.9	1657	16	US-10-147-496-433	Sequence 433, App
392	13	92.9	1657	16	US-10-145-630-433	Sequence 433, App	13	92.9	1657	16	US-10-147-505-433	Sequence 433, App
393	13	92.9	1657	16	US-10-145-747-433	Sequence 433, App	13	92.9	1657	16	US-10-147-516-433	Sequence 433, App
394	13	92.9	1657	16	US-10-145-752-433	Sequence 433, App	13	92.9	1657	16	US-10-152-398-433	Sequence 433, App
395	13	92.9	1657	16	US-10-145-754-433	Sequence 433, App	13	92.9	1657	16	US-10-139-980-433	Sequence 433, App
396	13	92.9	1657	16	US-10-145-755-433	Sequence 433, App	13	92.9	1657	16	US-10-145-750-433	Sequence 433, App
397	13	92.9	1657	16	US-10-145-818-433	Sequence 433, App	13	92.9	1657	16	US-10-152-373-433	Sequence 433, App
398	13	92.9	1657	16	US-10-145-820-433	Sequence 433, App	13	92.9	1657	16	US-10-121-044-433	Sequence 433, App
399	13	92.9	1657	16	US-10-145-872-433	Sequence 433, App	13	92.9	1657	16	US-10-121-055-433	Sequence 433, App
400	13	92.9	1657	16	US-10-145-873-433	Sequence 433, App	13	92.9	1657	16	US-10-121-057-433	Sequence 433, App
401	13	92.9	1657	16	US-10-147-481-433	Sequence 433, App	13	92.9	1657	16	US-10-121-058-433	Sequence 433, App
402	13	92.9	1657	16	US-10-147-482-433	Sequence 433, App	13	92.9	1657	16	US-10-121-059-433	Sequence 433, App
403	13	92.9	1657	16	US-10-147-503-433	Sequence 433, App	13	92.9	1657	16	US-10-121-060-433	Sequence 433, App
404	13	92.9	1657	16	US-10-147-522-433	Sequence 433, App	13	92.9	1657	16	US-10-123-109-433	Sequence 433, App
405	13	92.9	1657	16	US-10-152-401-433	Sequence 433, App	13	92.9	1657	16	US-10-123-154-433	Sequence 433, App
406	13	92.9	1657	16	US-10-157-783-433	Sequence 433, App	13	92.9	1657	16	US-10-123-157-433	Sequence 433, App
407	13	92.9	1657	16	US-10-158-792-433	Sequence 433, App	13	92.9	1657	16	US-10-123-806-433	Sequence 433, App
408	13	92.9	1657	16	US-10-158-822-433	Sequence 433, App	13	92.9	1657	16	US-10-124-814-433	Sequence 433, App
409	13	92.9	1657	16	US-10-143-035-433	Sequence 433, App	13	92.9	1657	16	US-10-124-816-433	Sequence 433, App
410	13	92.9	1657	16	US-10-145-751-433	Sequence 433, App	13	92.9	1657	16	US-10-124-820-433	Sequence 433, App
411	13	92.9	1657	16	US-10-145-822-433	Sequence 433, App	13	92.9	1657	16	US-10-125-704-433	Sequence 433, App
412	13	92.9	1657	16	US-10-145-824-433	Sequence 433, App	13	92.9	1657	16	US-10-125-927-433	Sequence 433, App
413	13	92.9	1657	16	US-10-145-827-433	Sequence 433, App	13	92.9	1657	16	US-10-142-889-433	Sequence 433, App
414	13	92.9	1657	16	US-10-145-869-433	Sequence 433, App	13	92.9	1657	16	US-10-145-874-433	Sequence 433, App
415	13	92.9	1657	16	US-10-145-875-433	Sequence 433, App	13	92.9	1657	16	US-10-147-497-433	Sequence 433, App
416	13	92.9	1657	16	US-10-145-877-433	Sequence 433, App	13	92.9	1657	16	US-10-152-371-433	Sequence 433, App
417	13	92.9	1657	16	US-10-145-958-433	Sequence 433, App	13	92.9	1657	16	US-10-152-374-433	Sequence 433, App
418	13	92.9	1657	16	US-10-146-787-433	Sequence 433, App	13	92.9	1657	16	US-10-152-375-433	Sequence 433, App
419	13	92.9	1657	16	US-10-146-790-433	Sequence 433, App	13	92.9	1657	16	US-10-152-377-433	Sequence 433, App
420	13	92.9	1657	16	US-10-146-793-433	Sequence 433, App	13	92.9	1657	16	US-10-152-386-433	Sequence 433, App
421	13	92.9	1657	16	US-10-147-480-433	Sequence 433, App	13	92.9	1657	16	US-10-152-391-433	Sequence 433, App
422	13	92.9	1657	16	US-10-147-485-433	Sequence 433, App	13	92.9	1657	16	US-10-152-399-433	Sequence 433, App
423	13	92.9	1657	16	US-10-147-486-433	Sequence 433, App	13	92.9	1657	16	US-10-156-848-433	Sequence 433, App
424	13	92.9	1657	16	US-10-147-487-433	Sequence 433, App	13	92.9	1657	16	US-10-157-785-433	Sequence 433, App
425	13	92.9	1657	16	US-10-147-490-433	Sequence 433, App	13	92.9	1657	16	US-10-157-794-433	Sequence 433, App
426	13	92.9	1657	16	US-10-147-494-433	Sequence 433, App	13	92.9	1657	16	US-10-157-796-433	Sequence 433, App
427	13	92.9	1657	16	US-10-147-498-433	Sequence 433, App	13	92.9	1657	16	US-10-160-500-433	Sequence 433, App
428	13	92.9	1657	16	US-10-147-514-433	Sequence 433, App	13	92.9	1657	16		
429	13	92.9	1657	16	US-10-147-524-433	Sequence 433, App	13	92.9	1657	16		
430	13	92.9	1657	16	US-10-152-379-433	Sequence 433, App	13	92.9	1657	16		
431	13	92.9	1657	16	US-10-152-394-433	Sequence 433, App	13	92.9	1657	16		
432	13	92.9	1657	16	US-10-152-406-433	Sequence 433, App	13	92.9	1657	16		
433	13	92.9	1657	16	US-10-156-847-433	Sequence 433, App	13	92.9	1657	16		
434	13	92.9	1657	16	US-10-157-778-433	Sequence 433, App	13	92.9	1657	16		
435	13	92.9	1657	16	US-10-157-799-433	Sequence 433, App	13	92.9	1657	16		
436	13	92.9	1657	16	US-10-160-504-433	Sequence 433, App	13	92.9	1657	16		
437	13	92.9	1657	16	US-10-145-634-433	Sequence 433, App	13	92.9	1657	16		
438	13	92.9	1657	16	US-10-147-520-433	Sequence 433, App	13	92.9	1657	16		
439	13	92.9	1657	16	US-10-157-781-433	Sequence 433, App	13	92.9	1657	16		
440	13	92.9	1657	16	US-10-176-989-433	Sequence 433, App	13	92.9	1657	16		
441	13	92.9	1657	16	US-10-147-491-433	Sequence 433, App	13	92.9	1657	16		
442	13	92.9	1657	16	US-10-152-378-433	Sequence 433, App	13	92.9	1657	16		
443	13	92.9	1657	16	US-10-152-382-433	Sequence 433, App	13	92.9	1657	16		
444	13	92.9	1657	16	US-10-152-383-433	Sequence 433, App	13	92.9	1657	16		
445	13	92.9	1657	16	US-10-152-384-433	Sequence 433, App	13	92.9	1657	16		
446	13	92.9	1657	16	US-10-152-387-433	Sequence 433, App	13	92.9	1657	16		
447	13	92.9	1657	16	US-10-152-389-433	Sequence 433, App	13	92.9	1657	16		
448	13	92.9	1657	16	US-10-152-390-433	Sequence 433, App	13	92.9	1657	16		
449	13	92.9	1657	16	US-10-152-392-433	Sequence 433, App	13	92.9	1657	16		

ALIGNMENTS

RESULT 1

US-10-010-476-15
; Sequence 15, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402A1bert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; METHYLTRANSFERASE AND METHODS FOR USE THEREOF
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchand, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-010-476-15
Query Match 100.0%; Score 14; DB 15; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
Db 1 TTGCCCCGCCCTT 14

RESULT 2
US-10-010-476-14
; Sequence 14, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402Albert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; METHYLTRANSFERASE AND METHODS FOR USE THEREOF
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Weiter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-010-476-9
Query Match. 100.0%; Score 14; DB 15; Length 30;

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-010-476-14
Query Match 100.0%; Score 14; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
Db 5 TTGCCCCGCCCTT 18

RESULT 3
US-10-010-476-9/c
; Sequence 9, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402Albert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; METHYLTRANSFERASE AND METHODS FOR USE THEREOF
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Weiter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-010-476-9
Query Match. 100.0%; Score 14; DB 15; Length 30;
```

Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
Db 22 TTGCCCCGCCCTT 9

RESULT 4

US-10-010-476-10
; Sequence 10, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402Albert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/010,476
; FILING DATE: 07-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/485,071
; FILING DATE: 12-Jun-1998
; APPLICATION NUMBER: 60/057,411
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30794.30W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-010-476-10

Query Match 100.0%; Score 14; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
Db 9 TTGCCCCGCCCTT 22

RESULT 5

US-10-010-476-13
; Sequence 13, Application US/10010476
; Publication No. US20030114402A1
; GENERAL INFORMATION:
; APPLICANT: Reich, No. US20030114402Albert O.
; Flynn, James
; TITLE OF INVENTION: MODULATORS OF DNA CYTOSINE-5

NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 11150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/010,476
FILING DATE: 07-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/485,071
FILING DATE: 12-Jun-1998
APPLICATION NUMBER: 60/057,411
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Canady, Karen S
REGISTRATION NUMBER: 39,927
REFERENCE/DOCKET NUMBER: 30794.30W001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310 445-1140
TELEFAX: 310 445-9031
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-010-476-13

Query Match 100.0%; Score 14; DB 15; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 14
Db 19 TTGCCCCGCCCTT 32

RESULT 6

US-10-242-535A-565/c
; Sequence 565, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 565
; LENGTH: 110
; TYPE: DNA

```
; ORGANISM: Human
US-10-242-535A-565

Query Match          100.0%; Score 14; DB 17; Length 110;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
DB 70 TTGCCCCCGCCCTT 57

RESULT 7
US-10-085-783A-565/c
; Sequence 565, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 565
; LENGTH: 110
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-565

Query Match          100.0%; Score 14; DB 17; Length 110;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
DB 70 TTGCCCCCGCCCTT 57

RESULT 8
US-09-960-352-462
; Sequence 462, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 462
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 03-BOWMS1-002-Q1-E1-A3
US-09-960-352-462

Query Match          100.0%; Score 14; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
```

```
DB 51 TTGCCCCCGCCCTT 64

RESULT 9
US-09-803-719-2383/c
; Sequence 2383, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2383
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-719-2383

Query Match          100.0%; Score 14; DB 10; Length 326;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
DB 128 TTGCCCCCGCCCTT 115

RESULT 10
US-09-803-719-2387/c
; Sequence 2387, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
```

; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Bitgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2387
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-719-2387

Query Match 100.0%; Score 14; DB 10; Length 340;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
||| ||||| ||||| |||||

Db 128 TTGCCCCGCCCCCTT 115

RESULT 11
US-10-425-115-153894/c
; Sequence 153894, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 153894
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(394)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_71930C.1
US-10-425-115-153894

Query Match 100.0%; Score 14; DB 18; Length 394;
Best Local Similarity 100.0%; Pred. No. 7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
||| ||||| ||||| |||||

Db 100 TTGCCCCGCCCCCTT 87

RESULT 12
US-10-425-115-131927
; Sequence 131927, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 131927
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_517C.1
US-10-425-115-131927

Query Match 100.0%; Score 14; DB 18; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
||| ||||| ||||| |||||

Db 362 TTGCCCCGCCCCCTT 375

RESULT 13
US-10-424-599-140509/c
; Sequence 140509, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 140509
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9788C.1
US-10-424-599-140509

Query Match 100.0%; Score 14; DB 17; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
||| ||||| ||||| |||||

Db 222 TTGCCCCGCCCCCTT 209

RESULT 14
US-10-437-963-23923
; Sequence 23923, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; OTHER INFORMATION: Plants and Uses Thereof for Plant Improvement


```
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 23923
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(600)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_28956C.1
US-10-437-963-23923

Query Match
Best Local Similarity 100.0%; Score 14; DB 18; Length 600;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 503 TTGCCCCGCCCTT 516

RESULT 15
US-10-437-963-49030/c
; Sequence 49030, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 49030
; LENGTH: 2352
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_5164C.1
US-10-437-963-49030

Query Match
Best Local Similarity 100.0%; Score 14; DB 18; Length 2352;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 785 TTGCCCCGCCCTT 772

RESULT 16
US-10-152-319A-2018
; Sequence 2018, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling

; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2018
; LENGTH: 2573
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_053618
US-10-152-319A-2018

Query Match
Best Local Similarity 100.0%; Score 14; DB 17; Length 2573;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 490 TTGCCCCGCCCTT 503

RESULT 17
US-10-133-937-26
; Sequence 26, Application US/10133937
; Publication No. US20030207278A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
; TITLE OF INVENTION: DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
; TITLE OF INVENTION: OTHER BIOLOGICAL STATES
; FILE REFERENCE: 11613.56US01
; CURRENT APPLICATION NUMBER: US/10/133,937
; CURRENT FILING DATE: 2002-11-04
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-133-937-26

Query Match
Best Local Similarity 100.0%; Score 14; DB 17; Length 2850;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 622 TTGCCCCGCCCTT 635
```

```
RESULT 18
US-10-159-563-26
; Sequence 26, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-26

Query Match 100.0%; Score 14; DB 17; Length 2850;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCGCCCTT 14
Db 622 TTGCCCCGCGCCCTT 635

RESULT 19
US-10-276-645-1
; Sequence 1, Application US/10276645
; Publication No. US20040087784A1
; GENERAL INFORMATION:
; APPLICANT: BASF-LYNX Bioscience
; TITLE OF INVENTION: Neuronal serine threonine protein kinase
; FILE REFERENCE: PCT1389
; CURRENT APPLICATION NUMBER: US/10/276,645
; CURRENT FILING DATE: 2003-07-04
; PRIOR APPLICATION NUMBER: DE20001024171
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3170
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: polyA signal
; LOCATION: (3094)..(3099)
; FEATURE:
; OTHER INFORMATION: Coding sequence: (1)...(2172)
US-10-276-645-1

Query Match 100.0%; Score 14; DB 17; Length 3170;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCGCCCTT 14
Db 2717 TTGCCCCGCGCCCTT 2730

RESULT 20
US-10-276-645-2
; Sequence 2, Application US/10276645
; Publication No. US20040087784A1
```

```
; GENERAL INFORMATION:
; APPLICANT: BASF-LYNX Bioscience
; TITLE OF INVENTION: Neuronal serine threonine protein kinase
; FILE REFERENCE: PCT1389
; CURRENT APPLICATION NUMBER: US/10/276,645
; CURRENT FILING DATE: 2003-07-04
; PRIOR APPLICATION NUMBER: DE20001024171
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3250
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: polyA signal
; LOCATION: (3174)..(3179)
; FEATURE:
; OTHER INFORMATION: Coding sequence: (1)...(1980)
US-10-276-645-2

Query Match 100.0%; Score 14; DB 17; Length 3250;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCGCCCTT 14
Db 2797 TTGCCCCGCGCCCTT 2810

RESULT 21
US-09-764-869-2349
; Sequence 2349, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2349
; LENGTH: 13273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-2349

Query Match 100.0%; Score 14; DB 9; Length 13273;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCGCCCTT 14
Db 9548 TTGCCCCGCGCCCTT 9561

RESULT 22
US-10-091-504-2349
; Sequence 2349, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2349
; LENGTH: 13273
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-2349

Query Match      100.0%; Score 14; DB 14; Length 13273;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
   |||||
Db 9548 TTGCCCCGCCCCCTT 9561

RESULT 23
US-10-227-577-2349
; Sequence 2349, Application US/10227577
; Publication No. US20040005575A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C2
; CURRENT APPLICATION NUMBER: US/10/227,577
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 10/091,504
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2349
; LENGTH: 13273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-227-577-2349

Query Match      100.0%; Score 14; DB 17; Length 13273;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
   |||||
Db 9548 TTGCCCCGCCCCCTT 9561

RESULT 24
US-10-189-268-11/c
; Sequence 11, Application US/10189268
; Publication No. US20040005570A1
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean
; APPLICANT: C. Frank Bennett
; APPLICANT: Kenneth W. Dobbie
; TITLE OF INVENTION: ANTISENSE MODULATION OF GERANYLGERANYL DIPHOSPHATE SYNTHASE 1 EXP
; FILE REFERENCE: PUS-0021
; CURRENT APPLICATION NUMBER: US/10/189,268
; CURRENT FILING DATE: 2002-07-02

; NUMBER OF SEQ ID NOS: 131
; SEQ ID NO 11
; LENGTH: 51001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1331-1430, 34714-34813
; OTHER INFORMATION: n = A,T,C or G
US-10-189-268-11

Query Match      100.0%; Score 14; DB 17; Length 51001;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
   |||||
Db 1912 TTGCCCCGCCCCCTT 1899

RESULT 25
US-10-741-600-20995
; Sequence 20995, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20995
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-20995

Query Match      92.9%; Score 13; DB 19; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCCCTT 14
   |||||
Db 43 TGCCCCGCCCCCTT 55

RESULT 26
US-10-741-600-21067
; Sequence 21067, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21067
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-21067

Query Match      92.9%; Score 13; DB 19; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCCCTT 14
   |||||
```

```
Db      84 TGCCCCGCCCTT 96

RESULT 27
US-10-741-600-21068
; Sequence 21068, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21068
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-21068

Query Match      92.9%; Score 13; DB 19; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCTT 14
      |||||
Db      46 TGCCCCGCCCTT 58

RESULT 28
US-10-369-493-25585/C
; Sequence 25585, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25585
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25585

Query Match      92.9%; Score 13; DB 17; Length 207;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCTT 14
      |||||
Db      194 TGCCCCGCCCTT 182

RESULT 29
US-10-477-797-2
; Sequence 2, Application US/10477797
; Publication No. US20040197308A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN SCIENCE AND TECHNOLOGY CORPORATION
; TITLE OF INVENTION: Replication Vector showing Cell-Specific Expression
; FILE REFERENCE: K01002PCT
; CURRENT APPLICATION NUMBER: US/10/477,797

; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: JP P2001-143999
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-477-797-2

Query Match      92.9%; Score 13; DB 18; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCTT 14
      |||||
Db      84 TGCCCCGCCCTT 96

RESULT 30
US-10-500-173-2
; Sequence 2, Application US/10500173
; Publication No. US20050032214A1
; GENERAL INFORMATION:
; APPLICANT: Katsuhito TAKAHASHI
; APPLICANT: Hisako YAMAMURA
; TITLE OF INVENTION: Cell specific expression/replication vector
; FILE REFERENCE: 4439-4022
; CURRENT APPLICATION NUMBER: US/10/500,173
; CURRENT FILING DATE: 2004-06-24
; PRIOR APPLICATION NUMBER: JP P2001-402102
; PRIOR FILING DATE: 2001-12-28
; PRIOR APPLICATION NUMBER: JP P2002-255395
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-500-173-2

Query Match      92.9%; Score 13; DB 19; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TGCCCCGCCCTT 14
      |||||
Db      84 TGCCCCGCCCTT 96

RESULT 31
US-09-923-876-326
; Sequence 326, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 326
; LENGTH: 265
; TYPE: DNA
```

; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700156758H1
; NAME/KEY: unsure
; LOCATION: 215, 220, 223-224, 226, 230
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-326

Query Match 92.9%; Score 13; DB 9; Length 265;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 89 TGCCCCGCCCTT 101

RESULT 32
US-09-923-876-326
; Sequence 326, Application US/09923876
; Publication No. US20030237110A9
; GENERAL INFORMATION:
; APPLICANT: Lalngudi, Rachunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 326
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030237110A9 700156758H1
; NAME/KEY: unsure
; LOCATION: 215, 220, 223-224, 226, 230
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-326

Query Match 92.9%; Score 13; DB 10; Length 265;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 89 TGCCCCGCCCTT 101

RESULT 33
US-10-425-115-29238
; Sequence 29238, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 29238

; LENGTH: 291
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(291)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_126674C.1
US-10-425-115-29238

Query Match 92.9%; Score 13; DB 18; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 131 TGCCCCGCCCTT 143

RESULT 34
US-10-425-115-60113/c
; Sequence 60113, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 60113
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154824C.1
US-10-425-115-60113

Query Match 92.9%; Score 13; DB 18; Length 332;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
|||
DB 111 TGCCCCGCCCTT 99

RESULT 35
US-10-477-797-3
; Sequence 3, Application US/10477797
; Publication No. US20040197308A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN SCIENCE AND TECHNOLOGY CORPORATION
; TITLE OF INVENTION: Replication Vector Showing Cell-Specific Expression
; FILE REFERENCE: K01002PCT
; CURRENT APPLICATION NUMBER: US/10/477,797
; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: JP P2001-143999
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Region consist
; OTHER INFORMATION: of human calponin gene promoter and its structural

; OTHER INFORMATION: gene fragment
US-10-477-797-3

Query Match 92.9%; Score 13; DB 18; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGCCCGGCCCTT 14
|||||
Db 84 TGGCCCGGCCCTT 96

RESULT 36

US-10-500-173-3
; Sequence 3, Application US/10500173
; Publication No. US2005003221A1
; GENERAL INFORMATION:
; APPLICANT: Katsuhito TAKAHASHI
; APPLICANT: Hisako YAMAMURA
; TITLE OF INVENTION: Cell specific expression/replication vector
; FILE REFERENCE: 4439-4022
; CURRENT APPLICATION NUMBER: US/10/500,173
; CURRENT FILING DATE: 2004-06-24
; PRIOR APPLICATION NUMBER: JP P2001-402102
; PRIOR FILING DATE: 2001-12-28
; PRIOR APPLICATION NUMBER: JP P2002-255395
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Region consist
; OTHER INFORMATION: of human calponin gene promoter and its structural
; OTHER INFORMATION: gene fragment
US-10-500-173-3

Query Match 92.9%; Score 13; DB 19; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGCCCGGCCCTT 14
|||||
Db 84 TGGCCCGGCCCTT 96

RESULT 37

US-10-425-115-159870/c
; Sequence 159870, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 159870
; LENGTH: 347
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_77375C.1
US-10-425-115-159870

Query Match 92.9%; Score 13; DB 18; Length 347;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 13
|||||
Db 29 TTGCCCCGCCCTT 17

RESULT 38

US-10-425-115-181957/c
; Sequence 181957, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 181957
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_97523C.1
US-10-425-115-181957

Query Match 92.9%; Score 13; DB 18; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 13
|||||
Db 30 TTGCCCCGCCCTT 18

RESULT 39

US-10-425-115-177239/c
; Sequence 177239, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 177239
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_93230C.1
US-10-425-115-177239

Query Match 92.9%; Score 13; DB 18; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGCCCGGCCCTT 14
|||||
Db 82 TGGCCCGGCCCTT 70

RESULT 40

US-10-116-802-513/c

```
; Sequence 513, Application US/10115802
; Publication No. US20030065157A1
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 513
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 902956.2
US-10-116-802-513

Query Match          92.9%; Score 13; DB 14; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
Db 340 TGCCCCGCCCTT 328

RESULT 41
US-09-918-995-21483
; Sequence 21483, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21483
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(467)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-21483

Query Match          92.9%; Score 13; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
Db 4 TTGCCCCGCCCTT 16

RESULT 42
US-10-767-701-5731/c
; Sequence 5731, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 5731
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2274_1
US-10-767-701-5731

Query Match          92.9%; Score 13; DB 18; Length 475;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGCCCCGCCCTT 14
Db 323 TGCCCCGCCCTT 311

RESULT 43
US-10-424-599-24118/c
; Sequence 24118, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 24118
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(477)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_12177C.1
US-10-424-599-24118

Query Match          92.9%; Score 13; DB 17; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 13
Db 107 TTGCCCCGCCCTT 95

RESULT 44
US-10-723-860-20/c
; Sequence 20, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
```

```
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-20

Query Match          92.9%; Score 13; DB 18; Length 529;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  2  TTGCCCCGCCCTT 14
      |||||
Db   212 TGCCCCGCCCTT 200

RESULT 45
US-10-029-386-12456/c
; Sequence 12456, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AEOWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12456
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: P33485, EVALUE 2.30e-02
; OTHER INFORMATION: NT HIT: G11431277, EVALUE 7.00e-62
; OTHER INFORMATION: EST_HUMAN HIT: BE061171.1, EVALUE 3.00e-12
US-10-029-386-12456

Query Match          92.9%; Score 13; DB 16; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  TTGCCCCGCCCTT 13
      |||||
Db   210 TTGCCCCGCCCTT 198

RESULT 46
US-10-152-319A-867/c
; Sequence 867, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Casle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
```

```
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 867
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AI102023
US-10-152-319A-867

Query Match          92.9%; Score 13; DB 17; Length 544;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  TTGCCCCGCCCTT 13
      |||||
Db   189 TTGCCCCGCCCTT 177

RESULT 47
US-10-425-115-96771/c
; Sequence 96771, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 96771
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_1976C.1
US-10-425-115-96771

Query Match          92.9%; Score 13; DB 18; Length 553;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  TTGCCCCGCCCTT 13
      |||||
Db   392 TTGCCCCGCCCTT 380

RESULT 48
US-10-767-701-16993
; Sequence 16993, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```


; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 16993
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(577)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3481-045-Q6-K1-C9
US-10-767-701-16993

Query Match 92.9%; Score 13; DB 18; Length 577;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 13
|||||
Db 478 TTGCCCCGCCCTT 490

RESULT 49

US-09-864-761-11958/c
; Sequence 11958, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11958
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL36172.12
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
US-09-864-761-11958

Query Match 92.9%; Score 13; DB 9; Length 581;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGCCCCGCCCTT 14
|||||
Db 240 TGCCCCGCCCTT 228

RESULT 50

US-10-425-115-10159/c
; Sequence 10159, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 10159
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_109269C.1
US-10-425-115-10159

Query Match 92.9%; Score 13; DB 18; Length 582;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGCCCCGCCCTT 13
|||||
Db 438 TTGCCCCGCCCTT 426

Search completed: April 25, 2005, 15:34:32
Job time : 178.759 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:07:14 ; Search time 295.448 Seconds

(without alignments)

2296.080 Million cell updates/sec

Title: US-10-010-476-15

Perfect score: 14

Sequence: 1 TTGCCCGGCCCTT 14

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description
	1	14	100.0	14	6	BD078131	BD078131 Modulator
	2	14	100.0	22	6	BD078130	BD078130 Modulator
C	3	14	100.0	30	6	BD078125	BD078125 Modulator
	4	14	100.0	30	6	BD078126	BD078126 Modulator
	5	14	100.0	50	6	BD078129	BD078129 Modulator
C	6	14	100.0	110	6	CQ55639	CQ55639 Sequence
	7	14	100.0	230	9	HS28B7R	255235 H. sapiens C
C	8	14	100.0	326	6	AX247453	AX247453 Sequence
	9	14	100.0	336	10	AF177024	AF177024 Mus muscu
C	10	14	100.0	340	6	AX247457	AX247457 Sequence
	11	14	100.0	411	6	AX070937	AX070937 Sequence
C	12	14	100.0	534	6	CQ735395	CQ735395 Sequence
C	13	14	100.0	611	8	AF183504	AF183504 Poiretia
C	14	14	100.0	612	8	AF183503	AF183503 Poiretia
	15	14	100.0	914	11	AF391824	AF391824 Sus scrofa
	16	14	100.0	1062	6	CQ576024	CQ576024 Sequence
C	17	14	100.0	1152	9	HS800176	HS800176 Homo sapi
	18	14	100.0	1200	10	MURNRF1	U70474 Mus musculu
C	19	14	100.0	1278	4	AF090321	AF090321 Bos tauru

C	20	14	100.0	1280	9	AF023265	AF023265 Homo sapi
C	21	14	100.0	1359	9	AF023266	AF023266 Homo sapi
C	22	14	100.0	1442	3	AF239667	AF239667 Drosophill
C	23	14	100.0	1475	4	AF090322	AF090322 Bos tauru
C	24	14	100.0	1515	6	CQ728469	CQ728469 Sequence
C	25	14	100.0	1548	9	BC001960	BC001960 Homo sapi
C	26	14	100.0	1590	9	BSM807538	BSM807538 Homo sapi
C	27	14	100.0	1597	9	HSU49283	HSU49283 Human NAD+-
C	28	14	100.0	1622	6	BD156852	BD156852 Primer fo
C	29	14	100.0	1622	6	AX877795	AX877795 Sequence
C	30	14	100.0	1632	9	AK001905	AK001905 Homo sapi
C	31	14	100.0	1632	10	AB020496S3	AB020496 Mus muscu
C	32	14	100.0	1704	10	MU5SYNDIG	MU5SYNDIG Mus muscu
C	33	14	100.0	1910	5	BC046660	BC046660 Xenopus l
C	34	14	100.0	1972	9	AF192789S1	AF192789 Homo sapi
C	35	14	100.0	2573	10	AF342738	AF342738 Rattus no
C	36	14	100.0	2850	9	AK055922	AK055922 Homo sapi
C	37	14	100.0	2881	10	AK122565	AK122565 Mus muscu
C	38	14	100.0	3170	6	AX305103	AX305103 Sequence
C	39	14	100.0	3250	6	AX305104	AX305104 Sequence
C	40	14	100.0	3665	6	CQ586766	CQ586766 Sequence
C	41	14	100.0	4226	6	CQ583298	CQ583298 Sequence
C	42	14	100.0	4878	6	CQ576023	CQ576023 Sequence
C	43	14	100.0	7782	5	XLJ67076	XLJ67076 Xenopus lae
C	44	14	100.0	23404	2	AC017160	AC017160 Drosophill
C	45	14	100.0	25666	2	AC020076	AC020076 Drosophill
C	46	14	100.0	26700	6	AR068044	AR068044 Sequence
C	47	14	100.0	26700	6	I91962	I91962 Sequence 1
C	48	14	100.0	26700	6	AR266365	AR266365 Sequence
C	49	14	100.0	28181	2	AC100717	AC100717 Mus muscu
C	50	14	100.0	30789	3	L13200	L13200 Caenorhabdi
C	51	14	100.0	33934	10	MMSYNDEIA	MMSYNDEIA 22532 M.musculu
C	52	14	100.0	37392	2	AC019617	AC019617 Drosophill
C	53	14	100.0	37451	2	AC015199	AC015199 Drosophill
C	54	14	100.0	37947	8	SCU39205	SCU39205 Saccharomyc
C	55	14	100.0	51050	3	AC004340	AC004340 Drosophill
C	56	14	100.0	60958	2	AC132203	AC132203 Homo sapi
C	57	14	100.0	68210	2	AC118623	AC118623 Mus muscu
C	58	14	100.0	69081	5	AC148409	AC148409 X. tropic
C	59	14	100.0	69165	2	AC023465	AC023465 Homo sapi
C	60	14	100.0	75354	2	AC142960_3	Continuation (4 of
C	61	14	100.0	76179	2	AC012976	AC012976 Drosophill
C	62	14	100.0	81735	2	AC022230	AC022230 Mus muscu
C	63	14	100.0	82261	9	AC000065	AC000065 Homo sapi
C	64	14	100.0	85539	2	AC019877	AC019877 Drosophill
C	65	14	100.0	87359	5	BX908397	BX908397 Zebrafish
C	66	14	100.0	89239	2	AC020126	AC020126 Drosophill
C	67	14	100.0	89387	2	AC015216	AC015216 Drosophill
C	68	14	100.0	110000	2	AC098252_0	Continuation (2 of
C	69	14	100.0	110000	2	AC102394_1	Continuation (8 of
C	70	14	100.0	110000	2	LMFLCHR31_07	Continuation (22 o
C	71	14	100.0	110000	2	LMFLCHR32_21	Continuation (13 o
C	72	14	100.0	110000	2	LMFLCHR36_12	Continuation (2 of
C	73	14	100.0	110000	3	AC125735_1	Continuation (3 of
C	74	14	100.0	110000	3	AC125735_2	Continuation (3 of
C	75	14	100.0	113285	10	AC018461	AC018461 Mus muscu
C	76	14	100.0	113887	10	BX539328	BX539328 Mouse DNA
C	77	14	100.0	118497	2	AC134479	AC134479 Rattus no
C	78	14	100.0	123663	1	AE017255	AE017255 Treponema
C	79	14	100.0	123628	9	AF000353	AF000353 Homo sapi
C	80	14	100.0	126322	2	AC008095	AC008095 Drosophill
C	81	14	100.0	128150	2	AC148748	AC148748 Sorex ara
C	82	14	100.0	129606	2	AC018486	AC018486 Drosophill
C	83	14	100.0	130802	10	AC142273	AC142273 Mus muscu
C	84	14	100.0	132671	10	AL732587	AL732587 Mouse DNA
C	85	14	100.0	136602	8	AF006168	AF006168 Oryza sat
C	86	14	100.0	137241	9	AC132219	AC132219 Homo sapi
C	87	14	100.0	137905	2	AC015605	AC015605 Mus muscu
C	88	14	100.0	138366	10	AC111733	AC111733 Rattus no
C	89	14	100.0	139891	10	AC090647	AC090647 Genomic a
C	90	14	100.0	142273	9	AL391994	AL391994 Human DNA
C	91	14	100.0	144581	2	AF004889	AF004889 Oryza sat
C	92	14	100.0	144907	2	AC131453	AC131453 Strongylio

c 93	14	100.0	148640	5	AF056116	AF056116 Fugu rubr	166	14	100.0	201868	2	AC149841	AC149841 Papio anu
94	14	100.0	148927	2	AC087127	AC087127 Mus muscu	167	14	100.0	206549	10	AL833794	AL833794 Mouse DNA
95	14	100.0	149478	5	BX571961	BX571961 Zebrafish	c 168	14	100.0	206665	2	AC092872	AC092872 Pan trogl
96	14	100.0	152444	9	AP000352	AP000352 Homo sapi	c 169	14	100.0	207198	2	CR376829	CR376829 Danio rer
c 97	14	100.0	152423	2	AC119192	AC119192 Mus muscu	170	14	100.0	207521	2	AC136980	AC136980 Mus muscu
c 98	14	100.0	152726	10	AC129195	AC129195 Mus muscu	c 171	14	100.0	207647	2	AC110352	AC110352 Rattus no
c 99	14	100.0	153436	2	AC115841	AC115841 Mus muscu	c 172	14	100.0	207818	9	AC0919206	AC0919206 Homo sapi
c 100	14	100.0	154336	3	AC008334	AC008334 Drosophil	c 173	14	100.0	208307	2	AC073351	AC073351 Mus muscu
c 101	14	100.0	154619	1	D90917	D90917 Synchocyst	c 174	14	100.0	209701	2	AC142179	AC142179 Rattus no
c 102	14	100.0	155521	9	AC069335	AC069335 Homo sapi	c 175	14	100.0	210220	2	AC149518	AC149518 Xenopus t
c 103	14	100.0	157345	10	AL772404	AL772404 Mouse DNA	176	14	100.0	210973	2	AC130949	AC130949 Rattus no
c 104	14	100.0	158443	10	MWHC214016	AF111102 Mus muscu	c 177	14	100.0	211550	9	AC073912	AC073912 Homo sapi
c 105	14	100.0	159243	2	AC135660	AC135660 Rattus no	c 178	14	100.0	212301	2	AC136735	AC136735 Mus muscu
c 106	14	100.0	159272	9	HSJ666C3	AL049712 Human DNA	c 179	14	100.0	212936	2	AC020970	AC020970 Mus muscu
c 107	14	100.0	160312	2	CR381603	CR381603 Danio rer	c 180	14	100.0	213263	10	AL593846	AL593846 Mouse DNA
c 108	14	100.0	160990	9	AL157702	AL157702 Human DNA	181	14	100.0	213688	3	AE003755	AE003755 Drosophil
c 109	14	100.0	161985	10	AL592169	AL592169 Mouse DNA	182	14	100.0	214839	2	AC124630	AC124630 Mus muscu
c 110	14	100.0	162612	2	AC149467	AC149467 Papio anu	183	14	100.0	214875	10	AC015584	AC015584 Mus muscu
c 111	14	100.0	162839	9	AL137846	AL137846 Human DNA	184	14	100.0	217290	2	AC074332	AC074332 Mus muscu
c 112	14	100.0	162876	3	AC008212	AC008212 Drosophil	185	14	100.0	217819	2	AC026386	AC026386 Mus muscu
c 113	14	100.0	162882	10	AC113003	AC113003 Mus muscu	186	14	100.0	218476	10	AC087116	AC087116 Mus muscu
c 114	14	100.0	162855	2	AC092877	AC092877 Homo sapi	187	14	100.0	218747	10	AC121599	AC121599 Mus muscu
c 115	14	100.0	163992	9	AC078889	AC078889 Homo sapi	188	14	100.0	219820	2	AC120135	AC120135 Mus muscu
c 116	14	100.0	164293	9	AC020978	AC020978 Homo sapi	189	14	100.0	219870	10	AL928644	AL928644 Mouse DNA
c 117	14	100.0	165474	2	AC025603	AC025603 Homo sapi	190	14	100.0	219957	2	AC147917	AC147917 Xenopus t
c 118	14	100.0	165519	3	AC008347	AC008347 Drosophil	c 191	14	100.0	220238	2	AC121710	AC121710 Rattus no
c 119	14	100.0	166000	3	AC104511	AC104511 Drosophil	192	14	100.0	220804	10	AC091283	AC091283 Mus muscu
c 120	14	100.0	166352	3	AC010669	AC010669 Drosophil	193	14	100.0	221050	2	AC097142	AC097142 Rattus no
c 121	14	100.0	166384	2	AC148889	AC148889 Otollemur	194	14	100.0	223952	10	AC107742	AC107742 Mus muscu
c 122	14	100.0	166513	2	AC073787	AC073787 Mus muscu	c 195	14	100.0	224003	10	AC140326	AC140326 Mus muscu
c 123	14	100.0	168612	9	AC108039	AC108039 Homo sapi	c 196	14	100.0	224607	2	AC108331	AC108331 Rattus no
c 124	14	100.0	169385	3	AC012164	AC012164 Drosophil	197	14	100.0	224818	10	AL845300	AL845300 Mouse DNA
c 125	14	100.0	170394	2	AC024371	AC024371 Homo sapi	c 198	14	100.0	226059	2	AC130572	AC130572 Rattus no
c 126	14	100.0	171041	2	AC148510	AC148510 Macropus	c 199	14	100.0	226594	10	AC087216	AC087216 Mus muscu
c 127	14	100.0	173672	2	AC121520	AC121520 Mus muscu	c 200	14	100.0	226673	2	AC099195	AC099195 Rattus no
c 128	14	100.0	174888	9	AC037459	AC037459 Homo sapi	201	14	100.0	227330	10	AC020786	AC020786 Mus muscu
c 129	14	100.0	175867	3	AC007976	AC007976 Drosophil	202	14	100.0	228131	2	AC127129	AC127129 Rattus no
c 130	14	100.0	175899	2	AC115205	AC115205 Rattus no	c 203	14	100.0	228935	2	AC074312	AC074312 Mus muscu
c 131	14	100.0	176184	2	AC119318	AC119318 Mus muscu	204	14	100.0	229196	2	AC103488	AC103488 Rattus no
c 132	14	100.0	177187	2	AC096848	AC096848 Papio anu	205	14	100.0	229491	2	AC124811	AC124811 Mus muscu
c 133	14	100.0	178040	2	AC138367	AC138367 Mus muscu	206	14	100.0	229494	10	AL627328	AL627328 Mouse DNA
c 134	14	100.0	178879	2	AC124052	AC124052 Mus muscu	207	14	100.0	229795	2	AC128336	AC128336 Rattus no
c 135	14	100.0	179537	3	AC013513	AC013513 Homo sapi	208	14	100.0	229860	2	AC139577	AC139577 Mus muscu
c 136	14	100.0	181053	3	AC022346	AC022346 Drosophil	209	14	100.0	230167	2	AC150046	AC150046 Gallus ga
c 137	14	100.0	181527	2	AC145576	AC145576 Mus muscu	c 210	14	100.0	230607	2	AC098023	AC098023 Rattus no
c 138	14	100.0	181636	2	AC123152	AC123152 Rattus no	211	14	100.0	232594	5	EX005006	EX005006 Zebrafish
c 139	14	100.0	181904	3	AC008211	AC008211 Drosophil	212	14	100.0	232634	2	AC099461	AC099461 Rattus no
c 140	14	100.0	182080	3	AC012373	AC012373 Drosophil	c 213	14	100.0	232634	2	AC099461	AC099461 Rattus no
c 141	14	100.0	182426	2	AC123443	AC123443 Rattus no	214	14	100.0	233869	2	AC134004	AC134004 Rattus no
c 142	14	100.0	182607	2	AC107814	AC107814 Mus muscu	c 215	14	100.0	234841	2	AC119696	AC119696 Rattus no
c 143	14	100.0	182685	2	LMPP881	AL590734 Leishmani	216	14	100.0	235554	2	AC096229	AC096229 Rattus no
c 144	14	100.0	182990	2	AC141383	AC141383 Rattus no	c 217	14	100.0	235956	2	AC098638	AC098638 Rattus no
c 145	14	100.0	183285	10	AC128665	AC128665 Mus muscu	218	14	100.0	235970	2	AC115388	AC115388 Rattus no
c 146	14	100.0	183440	2	AC069295	AC069295 Homo sapi	219	14	100.0	236454	2	AC130965	AC130965 Rattus no
c 147	14	100.0	184365	2	AC092639	AC092639 Mus muscu	220	14	100.0	236675	2	EX901904	EX901904 Danio rer
c 148	14	100.0	184650	3	AC092186	AC092186 Drosophil	221	14	100.0	237779	2	AC095321	AC095321 Rattus no
c 149	14	100.0	184675	9	AC145964	AC145964 Pan trogl	c 222	14	100.0	238602	2	AC109002	AC109002 Rattus no
c 150	14	100.0	187325	2	AC142432	AC142432 Rattus no	c 223	14	100.0	238671	2	AC125296	AC125296 Rattus no
c 151	14	100.0	189756	9	AC092327	AC092327 Homo sapi	c 224	14	100.0	239075	2	AC091259	AC091259 Mus muscu
c 152	14	100.0	190953	10	AC127286	AC127286 Mus muscu	c 225	14	100.0	239162	5	EX927077	EX927077 Zebrafish
c 153	14	100.0	191118	2	AC148890	AC148890 Otollemur	c 226	14	100.0	239328	2	AC123494	AC123494 Rattus no
c 154	14	100.0	192366	3	AC008369	AC008369 Drosophil	227	14	100.0	239416	2	AC130914	AC130914 Rattus no
c 155	14	100.0	193520	2	AC029986	AC029986 Rattus no	c 228	14	100.0	240262	2	AC129151	AC129151 Rattus no
c 156	14	100.0	193267	2	AC093695	AC093695 Homo sapi	229	14	100.0	240904	2	AC126204	AC126204 Rattus no
c 157	14	100.0	193765	2	AC118517	AC118517 Rattus no	c 230	14	100.0	241228	2	AC103429	AC103429 Rattus no
c 158	14	100.0	194129	2	AC147063	AC147063 Pan trogl	231	14	100.0	242135	2	AC127842	AC127842 Rattus no
c 159	14	100.0	196427	10	AC118017	AC118017 Mus muscu	232	14	100.0	242582	2	AC128397	AC128397 Rattus no
c 160	14	100.0	197005	2	AC137535	AC137535 Bos tauru	233	14	100.0	243210	2	AC136831	AC136831 Rattus no
c 161	14	100.0	197424	10	AC079042	AC079042 Mus muscu	234	14	100.0	243863	2	AC137319	AC137319 Rattus no
c 162	14	100.0	199574	2	AC022930	AC022930 Homo sapi	c 235	14	100.0	245076	2	AC129689	AC129689 Rattus no
c 163	14	100.0	201238	10	AL807395	AL807395 Mouse DNA	c 236	14	100.0	249058	2	AC121180	AC121180 Rattus no
c 164	14	100.0	201290	10	AL731682	AL731682 Mouse DNA	c 237	14	100.0	249360	2	AC127920	AC127920 Rattus no
c 165	14	100.0	201377	2	AC073767	AC073767 Mus muscu	c 238	14	100.0	250046	2	AC113592	AC113592 Mus muscu

c 239	14	100.0	250492	2	AC137179	AC137179 Rattus no	312	13	92.9	704	9	HS3441476	AJ341476 Homo sapi
c 240	14	100.0	253852	2	AC098562	AC098562 Rattus no	c 313	13	92.9	707	6	AX073110	AX073110 Sequence
c 241	14	100.0	253914	2	AC108617	AC108617 Rattus no	314	13	92.9	726	11	BV209237	BV209237 FXYP06_383
c 242	14	100.0	254454	2	AC121447	AC121447 Rattus no	c 315	13	92.9	728	9	HS326066	AJ326066 Homo sapi
c 243	14	100.0	256764	3	AE003792	AE003792 Drosophila	316	13	92.9	729	9	HS328391	AJ328391 Homo sapi
c 244	14	100.0	260166	2	AC133228	AC133228 Rattus no	c 317	13	92.9	747	9	HS323604	AJ323604 Homo sapi
c 245	14	100.0	261688	2	AC111685	AC111685 Rattus no	318	13	92.9	794	6	AR5800	AJ5800 Sequence 45
c 246	14	100.0	261927	2	AC126891	AC126891 Rattus no	319	13	92.9	794	6	AR155293	AR155293 Sequence
c 247	14	100.0	261990	2	AC111513	AC111513 Rattus no	320	13	92.9	794	6	E65818	E65818 Genome DNA
c 248	14	100.0	265382	2	AC114439	AC114439 Rattus no	321	13	92.9	822	8	PUR242076	AJ242076 Pogonotiu
c 249	14	100.0	265861	2	AC069274	AC069274 Mus muscu	c 322	13	92.9	824	10	WMDH1	X07295 M.musculus
c 250	14	100.0	270972	2	AC132556	AC132556 Rattus no	323	13	92.9	829	8	PUR242078	AJ242078 Pogonotiu
c 251	14	100.0	271178	3	AE003810	AE003810 Drosophila	324	13	92.9	831	8	PUR242080	AJ242080 Pogonotiu
c 252	14	100.0	271451	2	AC098503	AC098503 Rattus no	325	13	92.9	832	8	PUR242079	AJ242079 Pogonotiu
c 253	14	100.0	275036	2	AC150059	AC150059 Gallus ga	326	13	92.9	876	9	HS337862	AJ337862 Homo sapi
c 254	14	100.0	277307	2	AC105149	AC105149 Rattus no	c 327	13	92.9	878	9	HS323603	AJ323603 Homo sapi
c 255	14	100.0	287501	2	AC098561	AC098561 Rattus no	328	13	92.9	915	6	CQ727890	CQ727890 Sequence
c 256	14	100.0	293398	2	AE003525	AE003525 Drosophila	329	13	92.9	969	6	AX660905	AX660905 Sequence
c 257	14	100.0	298900	1	AP005937	AP005937 Bradyrhiz	c 330	13	92.9	1015	10	RNMGSR	X06554 Rat mRNA fo
c 258	14	100.0	299868	3	AE003436	AE003436 Drosophila	331	13	92.9	1019	8	AK058882	AK058882 Oryza sat
c 259	14	100.0	303641	3	AE003583	AE003583 Drosophila	332	13	92.9	1049	8	D55708	D55708 Oryza sativ
c 260	14	100.0	304262	1	AE017005	AE017005 Bacillus	c 333	13	92.9	1104	6	CQ414423	CQ414423 Sequence
c 261	14	100.0	304447	2	CR354432	CR354432 Danio rer	334	13	92.9	1106	8	AK073800	AK073800 Oryza sat
c 262	14	100.0	305109	3	AE003509	AE003509 Drosophila	c 335	13	92.9	1183	3	BMQPOU	L01266 Bombyx mori
c 263	14	100.0	332029	3	AE003491	AE003491 Drosophila	336	13	92.9	1195	6	AR321627	AR321627 Sequence
c 264	14	100.0	348134	1	BX640420	BX640420 Borgatell	c 337	13	92.9	1199	3	DROSINSUPB	Li6772 Drosophila
c 265	13	92.9	207	8	AY558120	AY558120 Saccharom	338	13	92.9	1221	9	D85611	D85611 Human smoot
c 266	13	92.9	211	8	AF183547	AF183547 Leucoside	c 339	13	92.9	1259	10	RNO312207	AJ312207 Rattus no
c 267	13	92.9	220	1	AF175753	AF175753 Unculture	340	13	92.9	1276	3	AK173563	AK173563 Clona int
c 268	13	92.9	240	9	HS97G9R	264138 H.sapiens C	c 341	13	92.9	1319	9	BC018415	BC018415 Homo sapi
c 269	13	92.9	260	6	BD181247	BD181247 Cell spec	342	13	92.9	1326	6	AR339376	AR339376 Sequence
c 270	13	92.9	260	6	BD182148	BD182148 Cell spec	c 343	13	92.9	1370	1	D87897	D87897 Serratia ma
c 271	13	92.9	279	9	HS8F4F	263771 H.sapiens C	c 344	13	92.9	1391	6	I07878	I07878 Serratia ma
c 272	13	92.9	314	9	HS12H4R	254537 H.sapiens C	c 345	13	92.9	1391	10	RATRSMRA	M11721 Rat brain-s
c 273	13	92.9	324	6	CQ744760	CQ744760 Sequence	c 346	13	92.9	1391	10	RNPEP2	M01544 Rat mRNA fr
c 274	13	92.9	333	6	BD181248	BD181248 Cell spec	c 347	13	92.9	1392	10	RATESNP	M35702 Rat brain-s
c 275	13	92.9	333	6	BD182149	BD182149 Cell spec	c 348	13	92.9	1398	6	I00254	I00254 Sequence 5
c 276	13	92.9	336	6	CQ746130	CQ746130 Sequence	c 349	13	92.9	1406	10	BC020003	BC020003 Mus muscu
c 277	13	92.9	387	10	AF188010	AF188010 Mus muscu	c 350	13	92.9	1409	6	AR363949	AR363949 Sequence
c 278	13	92.9	396	10	MMU17510	MMU17510 Mus muscu	c 351	13	92.9	1448	5	BC067721	BC067721 Danio rer
c 279	13	92.9	399	9	SCPA13	YJ28130 Homo sapi	c 352	13	92.9	1561	10	BC014809	BC014809 Mus muscu
c 280	13	92.9	420	8	SCPA13	X6050 S.cerevisia	c 353	13	92.9	1567	6	CQ850713	CQ850713 Sequence
c 281	13	92.9	420	10	MMPRN1	YJ9932 M.musculus	354	13	92.9	1567	9	AK127884	AK127884 Homo sapi
c 282	13	92.9	426	1	REMB013R2	AF116366 Rhizobium	c 355	13	92.9	1586	8	AJ620469	AJ620469 Euglena g
c 283	13	92.9	456	11	BV056070	BV056070 S212P6788	356	13	92.9	1598	5	CR848553	CR848553 Xenopus t
c 284	13	92.9	483	6	AX409686	AX409686 Sequence	c 357	13	92.9	1621	10	BC029017	BC029017 Mus muscu
c 285	13	92.9	483	9	HUMICAM1	M32331 Homo sapien	c 358	13	92.9	1628	5	BC062511	BC062511 Xenopus t
c 286	13	92.9	509	6	AR501521	AR501521 Sequence	c 359	13	92.9	1629	10	BC033396	BC033396 Mus muscu
c 287	13	92.9	509	6	AR516803	AR516803 Sequence	c 360	13	92.9	1631	10	BC033348	BC033348 Mus muscu
c 288	13	92.9	511	11	BV105892	BV105892 MARC 8233	c 361	13	92.9	1636	10	BC033423	BC033423 Mus muscu
c 289	13	92.9	523	8	AF523854	AF523854 Quercus r	c 362	13	92.9	1637	10	BC033383	BC033383 Mus muscu
c 290	13	92.9	524	9	AB088475	AB088475 Homo sapi	c 363	13	92.9	1638	10	BC055901	BC055901 Mus muscu
c 291	13	92.9	525	9	HS338219	AJ338219 Homo sapi	c 364	13	92.9	1639	10	BC034331	BC034331 Mus muscu
c 292	13	92.9	529	9	AF188179	AF188179 Homo sapi	c 365	13	92.9	1642	10	BC029661	BC029661 Mus muscu
c 293	13	92.9	543	9	HS338218	AJ338218 Homo sapi	c 366	13	92.9	1642	10	BC034595	BC034595 Mus muscu
c 294	13	92.9	560	9	AF212826	AF212826 Homo sapi	c 367	13	92.9	1645	10	BC031367	BC031367 Mus muscu
c 295	13	92.9	581	6	CQ099433	CQ099433 Sequence	c 368	13	92.9	1645	10	BC033380	BC033380 Mus muscu
c 296	13	92.9	581	6	CQ138415	CQ138415 Sequence	369	13	92.9	1647	6	AX086290	AX086290 Sequence
c 297	13	92.9	581	6	CQ221840	CQ221840 Sequence	370	13	92.9	1657	6	AR252516	AR252516 Sequence
c 298	13	92.9	581	6	CQ259783	CQ259783 Sequence	371	13	92.9	1657	6	AR528668	AR528668 Sequence
c 299	13	92.9	581	6	CQ297599	CQ297599 Sequence	c 372	13	92.9	1657	6	AX403311	AX403311 Sequence
c 300	13	92.9	581	6	CQ333905	CQ333905 Sequence	373	13	92.9	1657	9	AX358976	AX358976 Homo sapi
c 301	13	92.9	609	6	AR414331	AR414331 Sequence	374	13	92.9	1657	10	BC018258	BC018258 Mus muscu
c 302	13	92.9	609	6	AX971165	AX971165 Sequence	c 375	13	92.9	1668	6	AX660144	AX660144 Sequence
c 303	13	92.9	609	6	BD109884	BD109884 EST and e	376	13	92.9	1668	10	AB024448	AB024448 Mus muscu
c 304	13	92.9	615	9	HS3339384	AJ3339384 Homo sapi	c 377	13	92.9	1679	10	BC018480	BC018480 Mus muscu
c 305	13	92.9	631	11	BV041082	BV041082 S212P6049	c 378	13	92.9	1681	10	BC031916	BC031916 Mus muscu
c 306	13	92.9	648	9	HS3339172	AJ3339172 Homo sapi	c 379	13	92.9	1681	10	EC033358	EC033358 Mus muscu
c 307	13	92.9	654	11	BV077697	BV077697 S212P6214	c 380	13	92.9	1694	1	EC033358	EC033358 Mus muscu
c 308	13	92.9	656	9	BC029446	BC029446 Homo sapi	c 381	13	92.9	1725	6	AX430335	AX430335 Sequence
c 309	13	92.9	680	9	HS3329582	AJ3329582 Homo sapi	c 382	13	92.9	1732	10	BC033364	BC033364 Mus muscu
c 310	13	92.9	697	9	HS332750	AJ332750 Homo sapi	c 383	13	92.9	1761	9	HS0804122	AL832811 Homo sapi
c 311	13	92.9	700	5	S66408	S66408 c-erbB=prot	384	13	92.9				

385	13	92.9	1849	3	TBU20781	U20781 Trypanosoma	458	13	92.9	17736	6	CQ612314	CQ612314 Sequence
386	13	92.9	1850	9	BD18652	BD18652 Homo sapi	459	13	92.9	18611	10	AF330048	AF330048 Mus muscu
387	13	92.9	1910	6	BD191325	BD191325 186 human	460	13	92.9	18943	2	AC020236	AC020236 Drosophil
388	13	92.9	1910	6	AX924835	AX924835 Sequence	c 411	13	92.9	19867	2	AC017263	AC017263 Drosophil
389	13	92.9	1945	6	BD191199	BD191199 186 human	c 462	13	92.9	21927	10	MMU251761	AJ251761 Mus muscu
390	13	92.9	1945	6	AX924709	AX924709 Sequence	c 463	13	92.9	22004	2	AC017362	AC017362 Drosophil
391	13	92.9	1951	1	AF339883	AF339883 Rhodobact	464	13	92.9	22029	9	AC003967	AC003967 Homo sapi
392	13	92.9	1993	9	AC055389	AC055389 Homo sapi	c 465	13	92.9	22843	2	AC015232	AC015232 Drosophil
393	13	92.9	2017	9	AB026547	AB026547 Homo sapi	466	13	92.9	23142	9	HUMGNOS48	D26607 Homo sapien
c 394	13	92.9	2021	6	AX746924	AX746924 Sequence	c 467	13	92.9	23285	5	AY289003S1	AY289003 Danio rer
c 395	13	92.9	2021	6	AX091345	AX091345 Homo sapi	468	13	92.9	25106	9	AF519768	AF519768 Homo sapi
c 396	13	92.9	2058	6	AR389081	AR389081 Sequence	469	13	92.9	25305	3	CEF33C8	Z69790 Caenorhabdi
397	13	92.9	2121	6	CQ602801	CQ602801 Sequence	c 470	13	92.9	25546	2	AC017886	AC017886 Drosophil
398	13	92.9	2123	5	BC042223	BC042223 Xenopus l	471	13	92.9	25681	2	AC012952	AC012952 Drosophil
c 399	13	92.9	2148	10	RNMAGLR	X05301 Rat mRNA fo	472	13	92.9	27315	6	CQ582716	CQ582716 Sequence
c 400	13	92.9	2468	10	RATMAGB	M22357 Rat 1B236/m	473	13	92.9	28535	9	HSPRWTNP2	Z46940 Homo sapien
c 401	13	92.9	2475	6	AX7000174	AX7000174 Sequence	c 474	13	92.9	30573	2	AC015027	AC015027 Drosophil
c 402	13	92.9	2475	10	RATMAGA	M16800 Rat 1B236/m	c 475	13	92.9	3328	6	CQ584474	CQ584474 Sequence
403	13	92.9	2540	6	AX747406	AX747406 Sequence	476	13	92.9	32661	8	SC8010	Z49808 S.cerevisia
404	13	92.9	2540	9	AC092198	AC092198 Homo sapi	477	13	92.9	33795	2	AC012777	AC012777 Drosophil
405	13	92.9	2591	10	BC050781	BC050781 Mus muscu	478	13	92.9	35049	9	AV129569	AV129569 Homo sapi
c 406	13	92.9	2593	5	BC079720	BC079720 Xenopus l	479	13	92.9	35202	2	AC138478	AC138478 Homo sapi
407	13	92.9	2693	3	AK112201	AK112201 Ciona int	480	13	92.9	35353	2	AC020055	AC020055 Drosophil
408	13	92.9	2710	8	YSCDDR48A	M36110 S.cerevisia	c 481	13	92.9	35749	9	AC007162	AC007162 Homo sapi
409	13	92.9	2847	10	AY495668	AY495668 Mus muscu	482	13	92.9	36127	2	AC148871	AC148871 Chlamydom
410	13	92.9	2898	10	BC070403	BC070403 Mus muscu	483	13	92.9	36885	9	AP000526	AP000526 Homo sapi
c 411	13	92.9	2902	6	CQ596321	CQ596321 Sequence	484	13	92.9	38418	10	MUSPRNPA	U29186 Mus musculu
c 412	13	92.9	2907	5	BC063239	BC063239 Danio rer	c 485	13	92.9	38461	9	AC009007	AC009007 Homo sapi
c 413	13	92.9	2917	9	BC054514	BC054514 Homo sapi	486	13	92.9	38671	3	CSC01G10	Z81030 Caenorhabdi
414	13	92.9	2940	8	ZMU09987	U09987 Zea mays FR	487	13	92.9	38888	3	U41543	AC005559 Homo sapi
415	13	92.9	2941	9	AF3316597	AF3316597 Homo sapi	c 488	13	92.9	38911	9	AC005559	AC005559 Homo sapi
c 416	13	92.9	2992	10	MUSH02KJK	M63790 Mus musculu	c 489	13	92.9	39332	9	HSN38E12	Z68287 Human DNA s
c 417	13	92.9	3056	5	VAAWOL	X84017 V.ammodytes	490	13	92.9	39301	2	AP000525	AP000525 Homo sapi
c 418	13	92.9	3105	14	AF3331499	AF3331499 Gallid he	491	13	92.9	39407	2	AC010648	AC010648 Homo sapi
c 419	13	92.9	3159	8	TANADINAT	XY7965 Triticum ae	492	13	92.9	39760	10	U29187	U29187 Mus musculu
420	13	92.9	3210	1	SHP276161	AY276161 Streptomy	493	13	92.9	40573	9	HSU15422	U15422 Homo sapien
c 421	13	92.9	3218	9	HUNRYR32	U48480 Human skele	494	13	92.9	40839	6	CQ600137	CQ600137 Sequence
c 422	13	92.9	3259	6	AX827823	AX827823 Sequence	495	13	92.9	41461	9	AC010646	AC010646 Homo sapi
c 423	13	92.9	3259	10	RNNGFR	X05137 Rat mRNA fo	c 496	13	92.9	44323	3	AF098992	AF098992 Caenorhab
424	13	92.9	3386	8	AY044844	AY044844 Tricium	c 497	13	92.9	46183	2	AC104048	AC104048 Homo sapi
425	13	92.9	3508	10	MMU52821	US2821 Mus musculu	498	13	92.9	46777	9	BX247883	BX247883 Human DNA
c 426	13	92.9	3530	6	AX337055	AX337055 Sequence	c 499	13	92.9	47038	2	AC100944	AC100944 Mus muscu
c 427	13	92.9	3530	6	AX337584	AX337584 Sequence	c 500	13	92.9	49122	2	AC108385	AC108385 Pan trogl
c 428	13	92.9	3530	6	AX411127	AX411127 Sequence							
c 429	13	92.9	3530	9	HSPR8	X67247 H.sapiens r							
c 430	13	92.9	3631	1	CXEXBTONB	Z95386 X.campestri							
c 431	13	92.9	3645	6	CQ490818	CQ490818 Sequence							
c 432	13	92.9	3645	6	CQ492795	CQ492795 Sequence							
c 433	13	92.9	3645	6	CQ496663	CQ496663 Sequence							
c 434	13	92.9	3645	6	CQ497698	CQ497698 Sequence							
435	13	92.9	3840	4	AF251797	AF251797 Oryctolag							
436	13	92.9	4204	5	VAAWMC	X76731 V.ammodytes							
437	13	92.9	4597	9	BC075556	BC075556 Xenopus t							
438	13	92.9	4642	10	AK122261	AK122261 Mus muscu							
c 439	13	92.9	5105	3	AF115456	AF115456 Drosophil							
c 440	13	92.9	5138	4	AF143503	AF143503 Canis fam							
c 441	13	92.9	5164	2	AC015246	AC015246 Drosophil							
442	13	92.9	5216	3	AY060400	AY060400 Drosophil							
443	13	92.9	6217	4	D88184	D88184 Equus cabal							
c 444	13	92.9	6938	10	AF133277S5	AF133280 Mus muscu							
c 445	13	92.9	7085	2	AC015260	AC015260 Drosophil							
446	13	92.9	7504	9	AB040943	AB040943 Homo sapi							
447	13	92.9	7786	9	AF307340	AF307340 Homo sapi							
448	13	92.9	8693	4	OCWAPG	X52564 O.cuniculus							
449	13	92.9	9490	4	AF091049	AF091049 Canis fam							
450	13	92.9	9990	6	AX201603	AX201603 Sequence							
451	13	92.9	9990	6	AX201616	AX201616 Sequence							
452	13	92.9	10514	1	AE009934	AE009934 Pyrobacul							
c 453	13	92.9	11076	4	AY427787	AY427787 Bos tauru							
c 454	13	92.9	12682	14	AJ632327	AJ632327 Cotesia c							
c 455	13	92.9	12700	9	HSXBVII1	X71937 H.sapiens X							
c 456	13	92.9	15001	10	AY480043	AY480043 Mus muscu							
457	13	92.9	15473	9	AY13045683	AY130458 Homo sapi							

ALIGNMENTS

RESULT 1	BD078131	14 bp	DNA	linear	PAT 27-AUG-2002
BD078131	Modulator of DNA cytosine-5 methyltransferase and method of using the same.				
ACCESSION	BD078131				
VERSION	BD078131.1	GI:22623734			
KEYWORDS	JP 2001514862-A/15.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 14)				
AUTHORS	Reich,N.O. and Flynn,J.				
TITLE	Modulator of DNA cytosine-5 methyltransferase and method of using the same				
JOURNAL	Patent: JP 2001514862-A 15 18-SEP-2001;				
COMMENT	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA				
	OS Unidentified				
	PN JP 2001514862-A/15				
	PD 18-SEP-2001				
	PR 12-JUN-1998 JP 2000508978				
	PF 29-AUG-1997 US 60/057411				
	PI NORBERT O REICH, JAMES FLYNN				
	PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00				
	CC Strandedness: Double;				

```

CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
FH Key Location/Qualifiers
FT Source 1..14
/organism='Unidentified'.
FEATURES
    source
        Location/Qualifiers
            1..14
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 14;
Best Local Similarity 100.0%; Pred. No. 9.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
Db 1 TTGCCCCGCCCTT 14
RESULT 2
BD078130
LOCUS BD078130 22 bp DNA linear PAT 27-AUG-2002
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
ACCESSION BD078130
VERSION BD078130.1 GI:22623733
KEYWORDS JP 2001514862-A/14.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 22)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 14 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/14
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT Source 1..22
/organism='Unidentified'.
FEATURES
    source
        Location/Qualifiers
            1..22
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
Db 5 TTGCCCCGCCCTT 18
RESULT 3
BD078125/c
LOCUS BD078125 30 bp DNA linear PAT 27-AUG-2002
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.

```

```

the same.
BD078125
LOCUS BD078125.1 GI:22623728
KEYWORDS JP 2001514862-A/9.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 9 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/9
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT Source 1..30
/organism='Unidentified'.
FEATURES
    source
        Location/Qualifiers
            1..30
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
Db 22 TTGCCCCGCCCTT 9
RESULT 4
BD078126
LOCUS BD078126 30 bp DNA linear PAT 27-AUG-2002
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
ACCESSION BD078126
VERSION BD078126.1 GI:22623729
KEYWORDS JP 2001514862-A/10.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT Patent: JP 2001514862-A 10 18-SEP-2001;
OS Unidentified
PN JP 2001514862-A/10
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH,JAMES FLYNN
PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT Source 1..30

```

```
FT          Location/Qualifiers
FEATURES   source
           1..30
           /organism="unidentified"
           /mol_type="genomic DNA"
           /db_xref="taxon:32644"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.8e+03; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db  9 TTGCCCCGCCCCCTT 22
    |||||

RESULT 5
BD078129
LOCUS      BD078129.1 GI:22623732
DEFINITION Modulator of DNA cytosine-5 methyltransferase and method of using
            the same.
ACCESSION  BD078129
VERSION    JP 2001514862-A/13.
KEYWORDS   JP 2001514862-A/13.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 50)
AUTHORS   Reich,N.O. and Flynn,J.
TITLE      Modulator of DNA cytosine-5 methyltransferase and method of using
            the same
JOURNAL    Patent: JP 2001514862-A 13 18-SEP-2001;
            THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT    OS Unidentified
           PN JP 2001514862-A/13
           PD 18-SEP-2001
           PF 12-JUN-1998 JP 2000508978
           PR 29-AUG-1997 US 60/057411
           PI NORBERT O REICH,JAMES FLYNN
           PC C12N9/10,C12Q1/48//C12N15/09,C12N15/00
           CC Strandedness: Double;
           CC Topology: Linear;
           CC Modulator of DNA cytosine-5 methyltransferase and method of
           using the same
           CC Key using the same
           FT source 1..50
           FT Location/Qualifiers
           FT /organism="unidentified"
           FT /mol_type="genomic DNA"
           FT /db_xref="taxon:32644"
FEATURES   source
           1..50
           /organism="unidentified"
           /mol_type="genomic DNA"
           /db_xref="taxon:32644"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.3e+03; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db  19 TTGCCCCGCCCCCTT 32
    |||||

RESULT 6
CQ655639/c
LOCUS      CQ655639
DEFINITION Sequence 565 from Patent WO02070737.
ACCESSION  CQ655639
VERSION    CQ655639.1 GI:42113879
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
```

```
REFERENCE 1
AUTHORS   Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE      Compositions and methods relating to osteoarthritis
JOURNAL    Patent: WO 02070737-A 565 12-SEP-2002;
            Chondrogene Inc. (CA)
FEATURES   Location/Qualifiers
           source
           1..110
           /organism="Homo sapiens"
           /mol_type="unassigned DNA"
           /db_xref="taxon:9606"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 110;
Best Local Similarity 100.0%; Pred. No. 7.6e+03; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db  70 TTGCCCCGCCCCCTT 57
    |||||

RESULT 7
HS28B7R
LOCUS      HS28B7R
DEFINITION H.sapiens CpG island DNA genomic MseI fragment, Clone 28b7, reverse
            read cp928b7.rtia.
ACCESSION  Z55235
VERSION    255235.1 GI:1021276
KEYWORDS   CpG island; genomic MseI fragment.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE      Purification of CpG islands using a methylated DNA binding column
JOURNAL    Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE    94282070
PUBMED     8012384
REFERENCE  2 (bases 1 to 230)
AUTHORS   MacDonald,M., Huckle,E., Wilkinson,P. and Mickle,M.G.
TITLE      Direct Submision
JOURNAL    Submitted (18-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
COMMENT    Vector: pGEM-5zf(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hgmp.mrc.ac.uk/ for details
            or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES   Location/Qualifiers
           source
           1..230
           /organism="Homo sapiens"
           /mol_type="genomic DNA"
           /db_xref="taxon:9606"
           /clone="28b7"
           /sex="male"
           /tissue_type="blood"
           /clone_lib="CGI-1"
           /dev_stage="adult"
ORIGIN
Query Match      100.0%; Score 14; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 7e+03; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db  185 TTGCCCCGCCCCCTT 198
    |||||

RESULT 8
```



```

AX247453/c
LOCUS AX247453 326 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 2383 from Patent WO0166753.
ACCESSION AX247453
VERSION AX247453.1 GI:15862127
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kasam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0166753-A 2383 13-SEP-2001;
FEATURES Chiron Corporation (US) ; Hyseq Inc. (US)
source Location/Qualifiers
1..326
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 326;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
DB 128 TTGCCCCGCCCTT 115

RESULT 9
AF177024
LOCUS AF177024 336 bp mRNA linear ROD 03-MAY-2000
DEFINITION Mus musculus molossinus unknown mRNA.
ACCESSION AF177024
VERSION AF177024.1 GI:5834314
KEYWORDS
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Lee,Y.J., Park,C.W., Hahn,Y., Park,J., Lee,J., Yun,J.H., Hyun,B.
and Chung,J.H.
TITLE Mit1/Lb9 and Copg2, new members of mouse imprinted genes closely
linked to Peg1/Mest(1)
JOURNAL FEBS Lett. 472 (2-3), 230-234 (2000)
MEDLINE 10788617
PubMed 20250810
REFERENCE 2 (bases 1 to 336)
AUTHORS Lee,Y.J., Hahn,Y., Park,J., Park,C., Hyun,B. and Chung,J.H.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1999) Department of Biological Sciences, Korea
Advanced Institute of Science and Technology, 373-1 Gusong-dong
Yusong-gu, Taejon 305-701, South Korea
FEATURES Location/Qualifiers
1..336
/organism="Mus musculus molossinus"
/mol_type="mRNA"
/strain="KJR/Msf"
/sub_species="molossinus"
/db_xref="taxon:57486"
94
variation /note="as compared to Mus musculus domesticus C57BL/6J"
/replace="t"

ORIGIN
Query Match 100.0%; Score 14; DB 10; Length 336;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
DB 229 TTGCCCCGCCCTT 242

RESULT 10
AX247457/c
LOCUS AX247457 340 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 2387 from Patent WO0166753.
ACCESSION AX247457
VERSION AX247457.1 GI:15862131
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kasam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0166753-A 2387 13-SEP-2001;
FEATURES Chiron Corporation (US) ; Hyseq Inc. (US)
source Location/Qualifiers
1..340
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
DB 128 TTGCCCCGCCCTT 115

RESULT 11
AX070937/c
LOCUS AX070937 411 bp DNA linear PAT 25-JAN-2001
DEFINITION Sequence 1409 from Patent WO0102568.
ACCESSION AX070937
VERSION AX070937.1 GI:12581210
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
Kasam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Lamson,G., Drmanac,R., Crkvenjakov,R., Drmanac,S., Dickson,M.,
Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Stache-Crain,B.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0102568-A 1409 11-JAN-2001;
FEATURES CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
source Location/Qualifiers
1..411
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 TTGCCCCGCCCTT 14
Db      148 TTGCCCCGCCCTT 135

RESULT 12
LOCUS   CQ735395/c          534 bp      DNA      linear      PAT 03-FEB-2004
DEFINITION Sequence 21329 from Patent WO02068579.
ACCESSION CQ735395
VERSION   CQ735395.1 GI:42327138
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE    Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 21329 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source      Location/Qualifiers
            1..534
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 534;
Best Local Similarity 100.0%; Pred. No. 6.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
Db      109 TTGCCCCGCCCTT 96

RESULT 13
LOCUS   AF183504/c          611 bp      DNA      linear      PLN 13-JUL-2001
DEFINITION Poiretia punctata small subunit ribosomal RNA gene, partial
sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene
and internal transcribed spacer 2, complete sequence; and large
subunit ribosomal RNA gene, partial sequence.
ACCESSION AF183504
VERSION   AF183504.1 GI:6120044
KEYWORDS
SOURCE   Poiretia punctata
ORGANISM Poiretia punctata
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B.B., Sprent,J.I., de
Lima,H.C. and Gasson,P.E.
TITLE    The dalbergioid legumes (Fabaceae): delimitation of a pantropical
monophyletic clade
JOURNAL Am. J. Bot. 88 (3), 503-533 (2001)
PUBMED  11250829
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    The dalbergioid legumes
JOURNAL Unpublished
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    Direct Submission
JOURNAL Submitted (08-SEP-1999) Plant Sciences, Montana State University,
Ag Bioscience Bldg, Bozeman, MT 59717, USA
FEATURES
source      Location/Qualifiers
            1..612
            /organism="Poiretia angustifolia"
            /mol_type="genomic DNA"
            /specimen_voucher="Fonseca et al. 1419 (MO)"
            /db_xref="taxon:105946"
            /country="Brazil"
            <1..6
            /product="small subunit ribosomal RNA"
            7..209
            /product="internal transcribed spacer 1"
            210..378
            /product="5.8S ribosomal RNA"
            379..593
            /product="large subunit ribosomal RNA"

QY      1 TTGCCCCGCCCTT 14
Db      425 TTGCCCCGCCCTT 412

RESULT 14
LOCUS   AF183503/c          612 bp      DNA      linear      PLN 13-JUL-2001
DEFINITION Poiretia angustifolia small subunit ribosomal RNA gene, partial
sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene
and internal transcribed spacer 2, complete sequence; and large
subunit ribosomal RNA gene, partial sequence.
ACCESSION AF183503
VERSION   AF183503.1 GI:6120043
KEYWORDS
SOURCE   Poiretia angustifolia
ORGANISM Poiretia angustifolia
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B.B., Sprent,J.I., de
Lima,H.C. and Gasson,P.E.
TITLE    The dalbergioid legumes (Fabaceae): delimitation of a pantropical
monophyletic clade
JOURNAL Am. J. Bot. 88 (3), 503-533 (2001)
PUBMED  11250829
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    The dalbergioid legumes
JOURNAL Unpublished
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    Direct Submission
JOURNAL Submitted (08-SEP-1999) Plant Sciences, Montana State University,
Ag Bioscience Bldg, Bozeman, MT 59717, USA
FEATURES
source      Location/Qualifiers
            1..612
            /organism="Poiretia angustifolia"
            /mol_type="genomic DNA"
            /specimen_voucher="Fonseca et al. 1419 (MO)"
            /db_xref="taxon:105946"
            /country="Brazil"
            <1..6
            /product="small subunit ribosomal RNA"
            7..209
            /product="internal transcribed spacer 1"
            210..378
            /product="5.8S ribosomal RNA"
            379..593
            /product="large subunit ribosomal RNA"

QY      1 TTGCCCCGCCCTT 14
Db      425 TTGCCCCGCCCTT 412

RESULT 14
LOCUS   AF183503/c          612 bp      DNA      linear      PLN 13-JUL-2001
DEFINITION Poiretia angustifolia small subunit ribosomal RNA gene, partial
sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene
and internal transcribed spacer 2, complete sequence; and large
subunit ribosomal RNA gene, partial sequence.
ACCESSION AF183503
VERSION   AF183503.1 GI:6120043
KEYWORDS
SOURCE   Poiretia angustifolia
ORGANISM Poiretia angustifolia
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B.B., Sprent,J.I., de
Lima,H.C. and Gasson,P.E.
TITLE    The dalbergioid legumes (Fabaceae): delimitation of a pantropical
monophyletic clade
JOURNAL Am. J. Bot. 88 (3), 503-533 (2001)
PUBMED  11250829
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    The dalbergioid legumes
JOURNAL Unpublished
REFERENCE
AUTHORS Lavin,M., Pennington,R.T., Klitgaard,B. and Sprent,J.
TITLE    Direct Submission
JOURNAL Submitted (08-SEP-1999) Plant Sciences, Montana State University,
Ag Bioscience Bldg, Bozeman, MT 59717, USA
FEATURES
source      Location/Qualifiers
            1..612
            /organism="Poiretia angustifolia"
            /mol_type="genomic DNA"
            /specimen_voucher="Fonseca et al. 1419 (MO)"
            /db_xref="taxon:105946"
            /country="Brazil"
            <1..6
            /product="small subunit ribosomal RNA"
            7..209
            /product="internal transcribed spacer 1"
            210..378
            /product="5.8S ribosomal RNA"
            379..593
            /product="large subunit ribosomal RNA"

```

```

rRNA
ORIGIN
    /product="internal transcribed spacer 2"
    594..>612
    /product="large subunit ribosomal RNA"

Query Match      100.0%; Score 14; DB 8; Length 612;
Best Local Similarity 100.0%; Pred. No. 6.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db 425 TTGCCCCGCCCCCTT 412

RESULT 15
AF391824          914 bp DNA linear STS 12-FEB-2002
LOCUS
DEFINITION      Sus scrofa chromosome 18 clone SY32, sequence tagged site.
ACCESSION      AF391824
VERSION        AF391824.1 GI:14532316
KEYWORDS
SOURCE
ORGANISM      Sus scrofa (pig)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE      1 (bases 1 to 914)
AUTHORS        Campbell,E.M., Fahrenkrug,S.C., Vallet,J.L., Smith,T.P. and Rohrer,G.A.
TITLE          An updated linkage and comparative map of porcine chromosome 18
JOURNAL        Anim. Genet. 32 (6), 375-379 (2001)
MEDLINE        21599055
PUBMED        11736809
REFERENCE      2 (bases 1 to 914)
AUTHORS        Campbell,E.M.G., Fahrenkrug,S.C., Vallet,J.L., Smith,T.P.L. and Rohrer,G.A.
TITLE          Direct Submission
JOURNAL        Submitted (14-JUN-2001) USDA, ARS, U.S. Meat Animal Research Center, P.O. Box 166, Clay Center, Nebraska 68933-0166, USA
FEATURES
    source
    1..914
    /organism="Sus scrofa"
    /mol_type="genomic DNA"
    /db_xref="taxon:9823"
    /chromosome="18"
    /clone="SY32"
    30..144
    primer_bind 30..50
    primer_bind complement(126..144)

STS
primer_bind 30..50
primer_bind complement(126..144)

ORIGIN
    /product="internal transcribed spacer 2"
    594..>612
    /product="large subunit ribosomal RNA"

Query Match      100.0%; Score 14; DB 8; Length 612;
Best Local Similarity 100.0%; Pred. No. 6.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db 425 TTGCCCCGCCCCCTT 412

RESULT 16
AF391824          1062 bp DNA linear PAT 02-FEB-2004
LOCUS
DEFINITION      Sequence 3782 from Patent WO0171042.
ACCESSION      CQ576024
VERSION        CQ576024.1 GI:41639329
KEYWORDS
SOURCE
ORGANISM      Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1
AUTHORS        Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.

```

```

TITLE      Detection kits, such as nucleic acid arrays, for detecting the
JOURNAL    expression of 10,000 or more Drosophila genes and uses thereof
PE Corporation (NY) (US)
FEATURES
    source
    1..1062
    /organism="Drosophila sp."
    /mol_type="unassigned DNA"
    /db_xref="taxon:7242"

ORIGIN
    Query Match      100.0%; Score 14; DB 6; Length 1062;
    Best Local Similarity 100.0%; Pred. No. 5.9e+03;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCCCTT 14
    |||||
Db 602 TTGCCCCGCCCCCTT 615

RESULT 17
HSM800176/c      1152 bp mRNA linear PRI 18-FEB-2000
LOCUS
DEFINITION      Homo sapiens mRNA; cDNA DKFZp586P0718 (from clone DKFZp586P0718);
partial cds.
ACCESSION      AL050094
VERSION        AL050094.1 GI:4884117
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1152)
AUTHORS        Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE          Direct Submission
JOURNAL        Submitted (10-MAR-1999) MIPS, Am Klopferapitz 18a, D-82152
Martinsried, GERMANY
COMMENT       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the CDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp586P0718) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
    Location/Qualifiers
    1..1152
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="RZPD:DKFZp586P0718"
    /db_xref="taxon:9606"
    /clone="DKFZp586P0718"
    /tissue_type="uterus"
    /clone_lib="586 (synonym: hutel). Vector pSport1; host
    DH10B; sites NotI + SalI/MluI"
    /dev_stage="adult"
    1..1152
    /gene="DKFZp586P0718"
    /c1..781
    /note="NAD+-specific isocitrate dehydrogenase beta
    precursor"
    /codon_start=2
    /product="hypothetical protein"
    /protein_id="CAB43266.1"
    /db_xref="GI:4884118"
    /db_xref="GOA:Q9UG99"
    /db_xref="UniProt/TREMBL:Q9UG99"
    /translation="RTLVRLRRKLDLPANVHVHKSPLPGYVTRHNNLDLVIREQTEGE
    YSSLEHSARGVIECLKIVTRAKSQRIAKFAFDYATKKGRGKVTAVHKANIMKLDGL
    FLQCCVEVAELYPKIKFTETMIDCCMOLVQNPFQFDVLPNLYGNLIIDNLAAELVG
    GAGVVPGESYSAEYAVFTGARHPFAQAVGRNIANPTAMLLSASNNLRHLNLEYHSSM

```



```

complete cds.
AF023265
VERSION
AP023265.1 GI:4103445
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1280)
AUTHORS
Kim.Y.O., Koh.H.J., Kim.S.H., Jo.S.H., Huh.T.L., Jeong.K.S.,
Lee.I.J., Song.B.J. and Huh.T.L.
TITLE
Identification and functional characterization of a novel,
tissue-specific NAD(+)-dependent isocitrate dehydrogenase beta
subunit isoform
J. Biol. Chem. 274 (52), 36866-36875 (1999)
subunit isoform
10601238
PUBMED
2 (bases 1 to 1280)
REFERENCE
Ko,H.J., Kim,Y.O. and Huh,T.L.
AUTHORS
Direct Submission
TITLE
Submitted (08-SEP-1997) Department of genetic engineering,
Kyungpook National University, 1370 Sankyukdong, Puk-gu, Taegu
702-701, Republic of Korea
JOURNAL
Location/Qualifiers
FEATURES
1..1280
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
80..1231
/codon_start=1
/product="NAD+-specific isocitrate dehydrogenase beta
subunit isoform A"
/protein_id="AAD09339.1"
/db_xref="GI:4103446"
/translation="MAVLSGVRLTRALVSAGNPGAWRLSTSAHAASRSQAEDVR
VSGSPFTMLPGDGVGPDELHAKVFAAAVPVEFOEHLSEVQNNASSEKLEQVLS
SMKNKVALIIGKHTPMYKGLASDYMLRRKLDLFANVVHVKSLPCYVTRNNLDL
VIIRQTEGYSLSHESARGVIECLIKVTRAKSORIAKAFDYATKKGKGTAVVHK
ANIWKGLGLFQCCVEAEVLPKIFETMIIDNCQMOLVONPYQFDVLPVMPNLGNI
IDNLAAGLVGGVGVGSGSAEYAVFETGAPHPFAQAVGRNTANPTAMLLSASNMLR
HLNLYHSSMTADAVKKVIVKGVKVRTSDMGVYATCHDFTFVIALPFP"
LOCUS
AF023266
VERSION
AP023266.1 GI:4103447
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1359)
AUTHORS
Ko,H.J., Kim,Y.O. and Huh,T.L.
TITLE
Direct Submission
JOURNAL
Submitted (08-SEP-1997) Department of genetic engineering,
Kyungpook National University, 1370 Sankyukdong, Puk-gu, Taegu
702-701, Republic of Korea
FEATURES
1..1359
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
80..1243
/codon_start=1
/product="NAD+-specific isocitrate dehydrogenase beta
subunit isoform B"
/protein_id="AAD09340.1"
/db_xref="GI:4103448"
/translation="MAVLSGVRLTRALVSAGNPGAWRLSTSAHAASRSQAEDVR
VSGSPFTMLPGDGVGPDELHAKVFAAAVPVEFOEHLSEVQNNASSEKLEQVLS
SMKNKVALIIGKHTPMYKGLASDYMLRRKLDLFANVVHVKSLPCYVTRNNLDL
VIIRQTEGYSLSHESARGVIECLIKVTRAKSORIAKAFDYATKKGKGTAVVHK
ANIWKGLGLFQCCVEAEVLPKIFETMIIDNCQMOLVONPYQFDVLPVMPNLGNI
IDNLAAGLVGGVGVGSGSAEYAVFETGAPHPFAQAVGRNTANPTAMLLSASNMLR
HLNLYHSSMTADAVKKVIVKGVKVRTSDMGVYATCHDFTFVIALPFP"
LOCUS
AF023267
VERSION
AP023267.1 GI:7230769
KEYWORDS
SOURCE
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1442)
AUTHORS
Chalvet,F. and Samson,M.-L.
TITLE
Characterization of fly strains permitting GAL4-directed expression
of found in neurons
Genes 34 (1-2), 71-73 (2002)
JOURNAL
MEDLINE
22326002
PUBMED
12324951
REFERENCE
2 (bases 1 to 1442)
AUTHORS
Samson,M.L. and Chalvet,F.
TITLE
found in neurons, a third member of the Drosophila elav gene
family, encodes a neuronal protein and interacts with elav
Mech. Dev. 120 (3), 373-383 (2003)
JOURNAL
MEDLINE
22480022
PUBMED
12591606
REFERENCE
3 (bases 1 to 1442)
AUTHORS
Samson,M.-L.
TITLE
Direct Submission
JOURNAL
Submitted (25-FEB-2000) Embryologie, Universite Paris Sud, Orsay
91405, France
FEATURES
Location/Qualifiers
1..1442
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1..1442
/gene="elav"
59..1129
/gene="elav"
/notes="similar to ELAV protein"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAF43091.1"
/db_xref="GI:7230770"

```

/translation="MTNAMDIVKGSANGSDGSDNDSRTNLI VNYLPQTWTQBMERS
LFSSTIGELSKLVDRKVSGLNVLPSALTALNPALQOQSISGCVFVNTVRAEDAKAV
NTLNGRLQNKVIVKVSYPSSSTKGANLVVSLPKNLSDPLEGMPFASPGKIITSR
ILCNISLKGSGVIRIQDQNEAERAIQELNGKTPKGYAEPITVKEFANNPSAKAQ
IAPPTATYLTQQAATRRLAGALPSAGRIRYSPLAGLLANSILPGNMTGSGWCIF
VYNLAFTEENVLMQFGFFGAVQSVKVRIDLTQTSKCKGFGFVNTWYDEAVAIQSL
NGYTLGNRVLQVSFKNTKTT"

ORIGIN
Query Match 100.0%; Score 14; DB 3; Length 1442;
Best Local Similarity 100.0%; Pred. No. 5.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
|||||
Db 714 TTGCCCCGCCCTT 727

RESULT 23
AF090322/c
LOCUS AF090322 1475 bp mRNA linear MAM 13-DEC-2001
DEFINITION Bos taurus NAD(+)-isocitrate dehydrogenase subunit 1 IDH1-B
precursor (IDH) mRNA, nuclear gene encoding mitochondrial protein,
complete cds.
ACCESSION AF090322
VERSION AF090322.1 GI:3643260
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 1475)
AUTHORS Rushbrook,J.I., Weiss,C., Zeng,Y. and Huang,J.
TITLE Bovine NAD(+)-dependent Isocitrate Dehydrogenase: Alternative
Splicing and Tissue Dependent Expression of Subunit 1 (Abstract
518-M)
JOURNAL (in) PROTEIN SCIENCES; 142;
PROTEIN SOCIETY (1998)

REFERENCE 2 (bases 1 to 1475)
AUTHORS Weiss,C., Zeng,Y., Huang,J., Sobocka,M.B. and Rushbrook,J.I.
TITLE Bovine NAD(+)-dependent Isocitrate dehydrogenase: alternative
splicing and tissue-dependent expression of subunit 1
JOURNAL Biochemistry 39 (7), 1807-1816 (2000)
MEDLINE 20143498
PubMed 10677231

REFERENCE 3 (bases 1 to 1475)
AUTHORS Weiss,C., Zeng,Y., Huang,J. and Rushbrook,J.I.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1998) Biochemistry, SUNY Health Science Center at
Brooklyn, 450 Clarkson Ave. Brooklyn, NY 11203, USA

FEATURES
source
1..1475
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
1..1475
/gene="IDH"
55..1212
/EC number="1.1.1.41"
/note="alternative splice product; beta subunit; mRNA
present in kidney and liver; see 5' genomic sequence in
GenBank Accession Number AF090324"
/codon_start=1
/product="NAD(+)-isocitrate dehydrogenase subunit 1 IDH1-B
precursor"
protein_id="AAC83167.1"
/db_xref="GI:3643261"

/translation="MAALSRVRLTRALVAAPNPGAWRSLSCTSTVAQSSRTQEDVR
VEGAFPTMLPGDGGPELMHAKVEVFAASVPVEFQHHLSVEQNMASEKLSQVLS
SMKENKVAIIQKIHTPMYKGLASYNMLRKLDFANVHVHVKSLPGYKTRHNLDL
VIIREQTGEYSLSLEHESARGVIECLKIIVTRTKSQRIAKFAFDYATKKGKGVAVHK

ANIMKJGDLFLQCEEVAEALYPKIKPEKMIIDNCMQLVQNPQYQFDVLMVNPNYGNI
IDNLAAGLVGGVGVPGESYSAYAVPETAARHPFAQAGRNIAPTAMLLISANNMLR
HLNLEHSSNMAEAVKVKIKVGRTRDMGGYSTTTDFIKSVIGHLHPYGG"
transit_peptide 55..153
/gene="IDH"
mat_peptide 154..1209
/gene="IDH"
/product="NAD(+)-isocitrate dehydrogenase subunit 1
IDH1-B"

ORIGIN
Query Match 100.0%; Score 14; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 5.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
|||||
Db 689 TTGCCCCGCCCTT 676

RESULT 24
CQ728469/c
LOCUS CQ728469 1515 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 14403 from Patent WO02068579.
ACCESSION CQ728469
VERSION CQ728469.1 GI:42297364
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 14403 06-SEP-2002;
PE Corporation (NY) (US)

FEATURES
source
1..1515
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 1515;
Best Local Similarity 100.0%; Pred. No. 5.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCCCCGCCCTT 14
|||||
Db 636 TTGCCCCGCCCTT 623

RESULT 25
BC001960/c
LOCUS BC001960 1548 bp mRNA linear PRI 29-JUN-2004
DEFINITION Homo sapiens isocitrate dehydrogenase 3 (NAD+) beta, transcript
variant 1, mRNA (cDNA clone MGC:903 IMAGE:3536921), complete cds.
ACCESSION BC001960
VERSION BC001960.1 GI:12805012
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1548)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J.J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,

Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kerteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

2 (bases 1 to 1548)
Straussberg, R.

Direct Submission
Submitted (29-JAN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Anara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lilla Prabh, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalls, Jeff Stott, Miranda Teai, George Yang, Jacques Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 8 Row: d Column: 13
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28178818.

FEATURES
source
Location/Qualifiers
1..1548
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:903 IMAGE:3536921"
/tissue_type="lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOT87"
1..1548
/gene="IDH3B"
/note="synonyms: H-IDHB, FLJ11043, MGC903"
/db_xref="LocusID:3420"
/db_xref="MIM:604526"
19..1176
/gene="IDH3B"
/codon_start=1
/product="isocitrate dehydrogenase 3, beta subunit, isoform a precursor"
/protein_id="AAH01960.1"
/db_xref="GI:12805013"
/db_xref="LocusID:3420"
/db_xref="MIM:604526"
/translation="NAALSGVRWLTALVSGNPGAWRLSTSAHAARSQAEDVR
SGKSPFVTLPGDGVGPELMHAEVFKAAAPVPVEFQHHLSVEQNMASEKIKQVLS
SMKENKVAIGKIHTPMEYKGBELASYDMRLRRKLDLFANVVHVKSLPGYMTNRHNLID

VIIBQTEGYSSLEHESARGVIECLKIVTRAKSORIAKFAFDYATKKGRGKVTAVHK
ANIMKGLFLQCCCEVAELPKPKFTETMIIDNCQMLQVNPYQFDVLVMDLGNIGNI
IDNLAAGLGVGGVPGESYAEYAVFETGARHPFAQVGRNIAPTLASNNMLR
HLNLEYTHSSMIADAVKKVIVKVRTRDMGGYSTTTDFIKSVIGHLOTGKS"

ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 1548;
Best Local Similarity 100.0%; Pred. No. 5.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
|||||
Db 653 TTGCCCCCGCCCTT 640

RESULT 26
HSM807538/c 1590 bp mRNA linear PRI 30-AUG-2003
LOCUS HSM807538 Homo sapiens mRNA; cDNA DKFZp686K18144 (from clone DKFZp686K18144).
DEFINITION Homo sapiens mRNA; cDNA DKFZp686K18144 (from clone DKFZp686K18144).
ACCESSION BX647393
VERSION BX647393.1 GI:34366421
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1590)
AUTHORS Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
CONSRM The German Human cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp686K18144) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
source
Location/Qualifiers
1..1590
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686K18144"
/tissue_type="human colon endothel primary cell culture"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
polyA_signal 1549..1554
polyA_site 1561

ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14
|||||
Db 688 TTGCCCCCGCCCTT 675

RESULT 27
HSU49283/c 1597 bp mRNA linear PRI 20-DEC-1999
LOCUS HSU49283 Human NAD+-specific isocitrate dehydrogenase beta subunit
DEFINITION precursor, mRNA, nuclear gene encoding mitochondrial protein, complete cds.
ACCESSION U49283
VERSION U49283.1 GI:2737885

KEYWORDS Homo sapiens (human)

SOURCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1597) Kim, Y.O., Koh, H.J., Kim, S.H., Jo, S.H., Huh, J.W., Jeong, K.S., AUTHORS Lee, I.J., Song, B.J., and Huh, T.L.

TITLE Identification and functional characterization of a novel, tissue-specific NAD(+) -dependent isocitrate dehydrogenase beta subunit isoform

JOURNAL J. Biol. Chem. 274 (52), 36866-36875 (1999)

MEDLINE 20069662

PubMed 10601238

REFERENCE 2 (bases 1 to 1597) Ko, H., Kim, Y., Park, H., Oh, I., Yeo, S., Hong, S., Chae, B., Kim, S., AUTHORS Son, B., Lee, Y., Yeo, H., and Huh, T.-L.

TITLE Mitochondrial NAD+-specific isocitrate dehydrogenase

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1597) Huh, T.-L.

AUTHORS Direct Submission

TITLE Submitted (10-FEB-1996) Tae-Lin Huh, Kyungpook National University, COLLEGE OF NATURAL SCIENCES, Genetic Engineering, 1370 Sankyuk-Dong, Pook-Ku, Taegu, 702-701, Korea

JOURNAL Location/Qualifiers

FEATURES

source

1. .1597

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="hidb"

/sex="male"

/tissue_type="heart"

/dev_stage="adult"

1. .79

80. .1237

/codon_start=1

/product="NAD+-specific isocitrate dehydrogenase beta precursor"

/protein_id="AAB94295.1"

/db_xref="GI:2737886"

/translations="MAVLGSGVRLTRALVSAGNPGAWRGISTAAAHASRSQAEDVR VSGSFPTMLPGDVGPELHVAKEVFKAAAVPEFQEHLSVEQMASEEKLEQLVS SKNENKVAIGIKHTPMWEYGLASDLMRLRKLDLPANVVHVKSIPGMYTRHNRLD VIITREOTGEYSLSHESARGVTECLKIVTRAKSORIAKFPDYATKKGKGVTAVHG ANIMKLGLGFLQCCBEVAELPKIFETWIDNCCMLQVQNPYQFDVLVMPNLNGI INDLAAGLVGGAGVPGESYSABYAFETGARHPFAQVAGNRNTANPTAMLLSASMLR HLNLHYSMSMIADAVKVKVKGVRTRDMGGYSTTTTDFTKSVIGHLQTKGS"

transit_peptide 80. .181

mat_peptide 182. .1234

/product="NAD+-specific isocitrate dehydrogenase beta subunit"

/function="decarboxylation of isocitrate to alpha-ketoglutarate"

1235. .1597

3' UTR

ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1597;

Best Local Similarity 100.0%; Pred. No. 5.6e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCCGCCCTT 14

|||||

Db 714 TTGCCCCCGCCCTT 701

RESULT 28

BD156852/c

LOCUS

DEFINITION

Primer for synthesizing full-length cDNA and use thereof.

ACCESSION

BD156852

VERSION

BD156852.1 GI:278622610

KEYWORDS

JP 2002191363-A/11695.

ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 1622;
 Best Local Similarity 100.0%; Pred. No. 5.6e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 DB 745 TTGCCCCGCCCTT 732

RESULT 30
 AK001905/c

LOCUS Homo sapiens cDNA FLJ11043 fis, clone PLACE1004437, highly similar
 to Human NAD+-specific isocitrate dehydrogenase beta subunit
 precursor, mRNA.

ACCESSION AK001905.1 GI:7023460
 VERSION oligo capping; fis (full insert sequence).
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
 Sekine, M., Oba, Y., Nishi, T., Shibahara, T., Tanaka, T.,
 Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
 Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
 Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
 Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
 Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
 Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
 Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
 Yosida, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
 Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
 Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
 Shiohata, N., Sano, S., Moriya, S., Momiya, H., Satoh, N., Takami, S.,
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
 Kumsagi, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
 Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
 Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
 Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
 Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
 Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
 Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamaehita, R.,
 Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
 Complete sequencing and characterization of 21,243 full-length
 human cDNAs
 Nat. Genet. 36 (1), 40-45 (2004)

JOURNAL
 PUBMED 14702039

REFERENCE
 AUTHORS Isogai, T., Ota, T., Hayaishi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sato, H., Sugano, S., Aotsuka, S.,
 Yoshikawa, Y., Matsunawa, H., Ishii, S., Kawai, Y., Saito, K.,
 Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and
 Sasaki, N.

TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE
 AUTHORS Isogai, T. and Otsuki, T.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of

FEATURES
 source
 1. .1622
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1004437"
 /tissue_type="placenta"
 /clone_lib="PLACE1"
 /note="cloning vector: pME18SFL3"
 misc_feature
 536. .641
 /note="This 106 bases segment is not present in U49283,
 BC001960 and AL050094, probably caused by alternative
 splicing."
 567. .1268
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91971.1"
 /db_xref="GI:7023461"
 /translation="MKGERWSSLPFPVPVPSCCFLTQSGRVIECLKIVTRAKSQ
 IAKFAFDVATKRGKVTAVKANIMKLGDLQCCCEVAVSLYPIKPIETIVDNC
 MOLVQNPQFQVLYVMPNLYGNIIDNLAAGLVGGAGVPGESYASAYAFETGARHPFA
 QAVGRNIANTPMLLSASMLRHLNLEYHSSMIADAVKVKIKVGRVTRDMGYSYTTT
 DFKSVIGHLQTKGS"

ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1622;
 Best Local Similarity 100.0%; Pred. No. 5.6e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
 DB 745 TTGCCCCGCCCTT 732

RESULT 31
 AB020496S3

LOCUS Mus musculus Vax2 gene for homeobox protein, exon 3 and complete
 cds.

ACCESSION AB020498
 VERSION AB020498.1 GI:4589694
 KEYWORDS Vax2; homeobox protein.
 SEGMENT 3 of 3
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 JOURNAL Ohsaki, K., Ishida, Y., Morimitsu, T., Kominami, R. and Takahashi, N.
 REFERENCE Expression of the Vax family homeobox genes suggests multiple roles
 in eye development
 Genes Cells (1999) In press
 2 (bases 1 to 1632)
 JOURNAL Takahashi, N., Ohsaki, K., Ishida, Y. and Morimistu, T.
 REFERENCE Direct Submission
 AUTHORS Submitted (25-NOV-1998) Kanae Ohsaki, Nara Institute of Science and
 TITLE Technology, Graduate School of Biological Sciences, Takayama
 JOURNAL 8916-5, Ikoma, Nara 630-0101, Japan
 (E-mail: k-ohsaki@bs.aist-nara.ac.jp, Tel: 81-743-72-5470,
 Fax: 81-743-72-5479)

FEATURES
 source
 1. .1632
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /chromosome="6"

International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Takamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2850)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES
source
Location/Qualifiers
1..2850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MESAN2000572"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN2"
/note="cloning vector: pME18SFL3-primary culture, normal mesangial cells"

ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 2850;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
|||||
Db 622 TTGCCCCGCCCTT 635

RESULT 37
AK122565
LOCUS AK122565 2881 bp mRNA linear ROD 22-JUN-2004
DEFINITION Mus musculus mRNA for mKIAA1860 protein.
ACCESSION AK122565
VERSION AK122565.2 GI:49022881
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S., Nakajima, D., Nagase, T., Ohara, O. and Koga, H.
Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries DNA Res. 10 (1), 35-48 (2003)
MEDLINE 22579291
PUBMED 12693553
REFERENCE 2 (bases 1 to 2881)
Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
Direct Submission
Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan (E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
On Jun 22, 2004 this sequence version replaced gi:28972861.
The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing: Kazusa DNA Research

Institute; cDNA library construction, clone selection and 5' - & 3'-end one pass sequencing.
FEATURES
source
Location/Qualifiers
1..2881
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mid04050"
/note="vector: modified pBC SK+"
1..2881
/gene="mKIAA1860"
1..1905
/gene="mKIAA1860"
/note="CDS is predicted by in silico analysis. Start codon is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIAA1860 protein"
/protein_id="BAC55847.2"
/db_xref="GI:49022882"
/translation="VKLFVEIETKLLVMEYASAGEVDFYLVSHGRMKEKARAKF
RQIVSAVHYCHQKNI VHRDLKAEMLLDAENIKIADFGSNEFTLGSKLDFTCSGPP
YAPDLFOGKYDGEVDIWSLGIYLYTVSGSLPFDGHNLKELRVLGRKYRVPFY
MSTDCESILREFLVLPKRCITLQIMKDKWINGVEGELKPYTEPEDFGDTKRIE
VMVGMGVTRIEIKALTNQKNEVTATYLLGRKTEEGDRCAGLALARVPASDRTT
NGTSSSGSSHNKQORASSYTHRQRHSDFCGSPAPLHPARSPSTGDTLKEKERM
PGRKASCSAVGSGRGLPPSSPMVSSAHNPNAKIEIPERRKDSSTPNLPPSMMTRRN
TYVCTERPGRPSLLPNKENSCTSRVPPASPSHSLAPSPGSRSLARGSTIRST
FHGQVDRDRAGSGGVQNGPPASPTLAHEAAPLPGRPRPTNLTKLSKLTTRR
VTDPERIGGPETVTSCHLPWKTETAPLLFPMSVKLTSSRPPALMAALRQATAAA
RCRCRQPPFLACLHGAGGPEPLSHFEVEVCQLPRGLRGVLFRRVAGTALAFRTL
VTRISNDEL"

ORIGIN
Query Match 100.0%; Score 14; DB 10; Length 2881;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGCCCCGCCCTT 14
|||||
Db 2447 TTGCCCCGCCCTT 2460

RESULT 38
AX305103
LOCUS AX305103 3170 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 1 from Patent WO0188108.
ACCESSION AX305103
VERSION AX305103.1 GI:17644754
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Schneider, A., Klausner, B., Fischer, A., Newrzella, D., Goetz, B., Rosenner, M., Eisenhardt, G., Kuner, R., Trutzel, A., Kammandel, B., Jomana Naim, S. and Schwaninger, M.
Neuronal serine-threonine protein kinase Patent: WO 0188108-A 1 22-NOV-2001;
BASF-LYNX Bioscience AG (DE)
FEATURES
source
Location/Qualifiers
1..3170
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"
3094..3099
/note="Kodierende Sequenz: (1)_(2172)"

polyA_signal
source
Location/Qualifiers
3094..3099
/note="Kodierende Sequenz: (1)_(2172)"

ORIGIN
Query Match 100.0%; Score 14; DB 6; Length 3170;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      2717 TTGCCCCGCCCTT 2730

RESULT 39
AX305104
LOCUS      AX305104      3250 bp      DNA      linear      PAT 11-DEC-2001
DEFINITION Sequence 2 from Patent WO0188108.
ACCESSION  AX305104
VERSION     AX305104.1  GI:17644755
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
REFERENCE   1
AUTHORS     Rosener, A., Klausner, B., Fischer, A., Newrzella, D., Goetz, B.,
            Rosen, M., Eisenhardt, G., Kuner, R., Trutzel, A., Kammandel, B.,
            Jomana Naim, S. and Schwaninger, M.
TITLE       Neuronal serine-threonine protein kinase
JOURNAL     Patent: WO 0188108-A 2 22-NOV-2001;
            BASF-LYNX Bioscience AG (DE)
FEATURES   source
            1..3250
            /organism="Mus musculus"
            /mol_type="unassigned DNA"
            /db_xref="taxon:10090"
polya_signal 3174..3179
            /note="Kodierende Sequenz: (1)___.(1980)"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 3250;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      2797 TTGCCCCGCCCTT 2810

RESULT 40
CQ586766
LOCUS      CQ586766      3665 bp      DNA      linear      PAT 02-FEB-2004
DEFINITION Sequence 14524 from Patent WO0171042.
ACCESSION  CQ586766
VERSION     CQ586766.1  GI:41646490
KEYWORDS
SOURCE      Drosophila sp.
ORGANISM    Drosophila sp.
REFERENCE   1
AUTHORS     Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
TITLE       Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL     Patent: WO 0171042-A 14524 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES   source
            1..3665
            /organism="Drosophila sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:7242"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 3665;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      2717 TTGCCCCGCCCTT 2730

RESULT 41
CQ583298
LOCUS      CQ583298      4226 bp      DNA      linear      PAT 02-FEB-2004
DEFINITION Sequence 11056 from Patent WO0171042.
ACCESSION  CQ583298
VERSION     CQ583298.1  GI:41644178
KEYWORDS
SOURCE      Drosophila sp.
ORGANISM    Drosophila sp.
REFERENCE   1
AUTHORS     Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
TITLE       Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL     Patent: WO 0171042-A 11056 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES   source
            1..4226
            /organism="Drosophila sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:7242"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 4226;
Best Local Similarity 100.0%; Pred. No. 5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      2736 TTGCCCCGCCCTT 2723

RESULT 42
CQ576023
LOCUS      CQ576023      4878 bp      DNA      linear      PAT 02-FEB-2004
DEFINITION Sequence 3781 from Patent WO0171042.
ACCESSION  CQ576023
VERSION     CQ576023.1  GI:41639328
KEYWORDS
SOURCE      Drosophila sp.
ORGANISM    Drosophila sp.
REFERENCE   1
AUTHORS     Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
TITLE       Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL     Patent: WO 0171042-A 3781 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES   source
            1..4878
            /organism="Drosophila sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:7242"
ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 4878;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCGCCCTT 14
      |||||
Db      2009 TTGCCCCGCCCTT 2022

RESULT 43
XLU67076

```

```

LOCUS       XLU67076               7782 bp    DNA        linear       VRT 03-OCT-1997
DEFINITION  Xenopus laevis zing finger protein XFDL156 gene, complete cds.
ACCESSION   U67076
VERSION     U67076.1  GI:1519543
KEYWORDS    .
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
              Xenopodinae; Xenopus; Xenopus.
REFERENCE   1 (bases 1 to 7782)
AUTHORS     Bellefroid,E., Bourguignon,C., Bouwmeester,T., Rausch,O.,
              Blumberg,B. and Pieler,T.
TITLE       Transcription regulation and alternative splicing of an early
              zygotic gene encoding two structurally distinct zinc finger
              proteins in Xenopus laevis
JOURNAL     Mech. Dev. 63 (1), 99-108 (1997)
MEDLINE     97321550
PUBMED      9178260
REFERENCE   2 (bases 1 to 7782)
AUTHORS     Bellefroid,E., Bourguignon,C., Bouwmeester,T., Rausch,O. and
              Pieler,T.
TITLE       Direct Submission
JOURNAL     Submitted (16-AUG-1996) Biochemistry, Göttingen University,
              Humboldtallee 23, Göttingen 37073, Germany
FEATURES    Location/Qualifiers
             source
               1..7782
               /organism="Xenopus laevis"
               /mol_type="genomic DNA"
               /db_xref="taxon:8355"
               join(2748..2861,4879..4984,5234..7062)
               join(4967..4984,5234..6937)
               /note="zinc finger protein"
               /codon_start=1
               /product="XFDL156"
               /protein_id="AAB72013.1"
               /db_xref="GI:1519544"
               /translation="MMHLISNWPKEKQNLNGSEMTRMKKILEVSGSERINWGNLSQT
               QTHKESDRTPNQYSPMPQCYNGKGFALQREKLHNTNVHVRMPNIAKQPKFNRE
               IGSDSPKRAKSSKSHVANIASHQTKINKESKAQSAVKPFCICIKRKFVARNLL
               THORLHTGRFRFYCKGKCFNKKIHIHAYVHTGKPPICTKCNKGFJWNRDLQEH
               KSKEMRNKSLGLVRHDKLREVSTDPCKREKSFCTCEGKSFKRSKLTHTFLCHTG
               EKPFCVCHGKGFDRNYKLSLHLRHTGENLSVCPDCGKSYTDKNKLIWMRLHTTEA
               FMCSEGRKSPFIYNLKSLQHTGKFPVCECGKSFIRKYEKFSLLHLIHTGKYPFI
               CSCEGKSPMDKRYLKHNSVHGKTFPCTCEGKFAAKNKRQHMHTGKPHKTE
               CGKLEKSLERHLSHTGKVPISCFCEGQFTWKHLIQHLSHTGKRPVCSGEG
               KSKTKASLTVCHHTGKFPVCECGKSYRSHGALRYHLQAHIKGKYFAHSEHGES
               YKLENG"

ORIGIN
Query Match      100.0%; Score 14; DB 5; Length 7782;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCCGCCCTT 14
        |||
        554 TTGCCCCCGCCCTT 567

RESULT 44
AC017160
LOCUS       AC017160               23404 bp    DNA        linear       HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017160
VERSION     AC017160.1  GI:6553826
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
              Adams,M. and Venter,J.C.
REFERENCE   1 (bases 1 to 23404)
AUTHORS     Adams,M. and Venter,J.C.

LOCUS       AC017160               23404 bp    DNA        linear       HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017160
VERSION     AC017160.1  GI:6553826
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
              Adams,M. and Venter,J.C.

```

```

Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10209707 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES    Location/Qualifiers
             source
               1..23404
               /organism="Drosophila melanogaster"
               /mol_type="genomic DNA"
               /db_xref="taxon:7227"

ORIGIN
Query Match      100.0%; Score 14; DB 2; Length 23404;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCCGCCCTT 14
        |||
        4897 TTGCCCCCGCCCTT 4910

RESULT 45
AC020076
LOCUS       AC020076               25666 bp    DNA        linear       HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020076
VERSION     AC020076.1  GI:6664821
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
              Adams,M. and Venter,J.C.
REFERENCE   1 (bases 1 to 25666)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
              Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211962 by the submitter.
              For more information on this record e-mail to fly@celera.com.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
FEATURES    Location/Qualifiers
             source
               1..25666
               /organism="Drosophila melanogaster"
               /mol_type="genomic DNA"
               /db_xref="taxon:7227"

ORIGIN
Query Match      100.0%; Score 14; DB 2; Length 25666;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGCCCCCGCCCTT 14
        |||
        6217 TTGCCCCCGCCCTT 6230

RESULT 46
AR068044/c
LOCUS       AR068044               26700 bp    DNA        linear       PAT 29-SEP-1999
DEFINITION  Sequence 5 from patent US 5851993.
ACCESSION   AR068044
VERSION     AR068044.1  GI:5999266
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.

```

```

Unclassified.
REFERENCE 1 (bases 1 to 26700)
AUTHORS Jalkanen,M. and Mali,M.
TITLE Suppression of tumor cell growth by syndecan-1 ectodomain
JOURNAL Patent: US 5851993-A 5 22-DEC-1998;
FEATURES
    source
        1. .26700
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 4027 TTGCCCCGCCCTT 4014

RESULT 47
LOCUS I91962/c
DEFINITION Sequence 1 from patent US 5726058.
ACCESSION I91962
VERSION I91962.1 GI:3936432
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 26700)
AUTHORS Jalkanen,M., Alanen-Kurki,L., Auvinen,P., Jaakkola,P., Leppa,S.,
        Mali,M., Vihinen,T. and Warri,A.
TITLE Syndecan stimulation of cellular differentiation
JOURNAL Patent: US 5726058-A 1 10-WAR-1998;
FEATURES
    source
        1. .26700
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 4027 TTGCCCCGCCCTT 4014

RESULT 48
LOCUS AR266365/c
DEFINITION Sequence 1 from patent US 6492344.
ACCESSION AR266365
VERSION AR266365.1 GI:29695298
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 26700)
AUTHORS Jalkanen,M., Jaakkola,P. and Vihinen,T.
TITLE Syndecan enhancer element and syndecan stimulation of cellular
        differentiation
JOURNAL Patent: US 6492344-A 1 10-DEC-2002;
FEATURES
    source
        1. .26700
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Query Match      100.0%; Score 14; DB 6; Length 26700;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
Db 4027 TTGCCCCGCCCTT 4014

RESULT 49
LOCUS AC100717/c
DEFINITION Mus musculus clone RP23-170B13, linear
ACCESSION AC100717
VERSION AC100717.1 GI:17048083
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 28181)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-170B13
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 28181)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
        Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
        Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
        Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
        Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
        Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
        Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
        Hages,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
        Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
        Lanazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
        Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
        McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
        Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
        Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
        Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
        Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
        Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
        Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
        Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
        Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
        Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
        Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
        Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15994
Center clone name: 170_B_13
-----
* NOTE: This record contains 34 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 742: contig of 742 bp in length
* 743 842: gap of 100 bp

```



```

* 843 1600: contig of 758 bp in length
* 1601 1700: gap of 100 bp
* 1701 2466: contig of 766 bp in length
* 2467 2567: gap of 100 bp
* 2567 3317: contig of 751 bp in length
* 3318 3417: gap of 100 bp
* 3418 4128: contig of 711 bp in length
* 4129 4228: gap of 100 bp
* 4229 4955: contig of 727 bp in length
* 4956 5055: gap of 100 bp
* 5056 5747: contig of 692 bp in length
* 5748 5847: gap of 100 bp
* 5848 6607: contig of 760 bp in length
* 6608 6707: gap of 100 bp
* 6708 7380: contig of 673 bp in length
* 7381 7480: gap of 100 bp
* 7481 8204: contig of 724 bp in length
* 8205 8304: gap of 100 bp
* 8305 9043: contig of 739 bp in length
* 9044 9443: gap of 100 bp
* 9443 9958: contig of 715 bp in length
* 9959 10764: contig of 806 bp in length
* 10765 10864: gap of 100 bp
* 10865 11508: contig of 744 bp in length
* 11509 11708: gap of 100 bp
* 11709 12442: contig of 734 bp in length
* 12443 12542: gap of 100 bp
* 12543 13237: contig of 695 bp in length
* 13238 13337: gap of 100 bp
* 13338 14008: contig of 671 bp in length
* 14009 14108: gap of 100 bp
* 14109 14850: contig of 742 bp in length
* 14851 14950: gap of 100 bp
* 14951 15568: contig of 718 bp in length
* 15569 15768: gap of 100 bp
* 15769 16483: contig of 715 bp in length
* 16484 16583: gap of 100 bp
* 16584 17339: contig of 756 bp in length
* 17340 17439: gap of 100 bp
* 17440 18186: contig of 747 bp in length
* 18187 18286: gap of 100 bp
* 18287 19007: contig of 721 bp in length
* 19008 19107: gap of 100 bp
* 19108 19788: contig of 681 bp in length
* 19789 19888: gap of 100 bp
* 19889 20624: contig of 736 bp in length
* 20625 20724: gap of 100 bp
* 20725 21418: contig of 694 bp in length
* 21419 21518: gap of 100 bp
* 21519 22280: contig of 762 bp in length
* 22281 22380: gap of 100 bp
* 22381 23127: contig of 747 bp in length
* 23128 23227: gap of 100 bp
* 23228 23956: contig of 729 bp in length
* 23957 24056: gap of 100 bp
* 24057 24780: contig of 724 bp in length
* 24781 24880: gap of 100 bp
* 24881 25508: contig of 728 bp in length
* 25509 25708: gap of 100 bp
* 25709 26525: contig of 817 bp in length
* 26526 26625: gap of 100 bp
* 26626 27380: contig of 755 bp in length
* 27381 27480: gap of 100 bp
* 27481 28181: contig of 701 bp in length.

```

FEATURES source

```

1..28181
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /clone="RP23-170B13"
  /clone_lib="RPC1-23 Female Mouse BAC"

```

ORIGIN

```

Query Match      100.0%; Score 14; DB 2; Length 28181;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCCGCCCTT 14
    |||||
Db 4116 TTGCCCCGCCCTT 4103

RESULT 50
L13200
LOCUS L13200 30789 bp DNA linear INV 29-SEP-2004
DEFINITION Caenorhabditis elegans cosmid ZK1236, complete sequence.
ACCESSION L13200
VERSION L13200.3 GI:15963744
KEYWORDS HTG.
SOURCE
  ORGANISM
    Caenorhabditis elegans
    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
    Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
  1 (bases 1 to 30789)
    WormBase Consortium
    Genome sequence of the nematode C. elegans: a platform for
    investigating biology. The C. elegans Sequencing Consortium
    Science 282 (5396), 2012-2018 (1998)
    99069613
    PUBMED
  2 (bases 1 to 30789)
    Favello,A.
    The sequence of C. elegans cosmid ZK1236
    Unpublished (2001)
  3 (bases 1 to 30789)
    Waterston,R.
    Direct Submission
    Submitted (11-JUL-2001) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  4 (bases 1 to 30789)
    Waterston,R.
    Direct Submission
    Submitted (05-OCT-2001) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  5 (bases 1 to 30789)
    Waterston,R.
    Direct Submission
    Submitted (29-MAY-2002) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  6 (bases 1 to 30789)
    Waterston,R.
    Direct Submission
    Submitted (22-NOV-2002) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  7 (bases 1 to 30789)
    Wilson,R.
    Direct Submission
    Submitted (03-DEC-2003) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  8 (bases 1 to 30789)
    WormBase Consortium
    Direct Submission
    Submitted (22-SEP-2004) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  9 (bases 1 to 30789)
    WormBase Consortium

```

TITLE
JOURNAL

Direct Submission
Submitted (29-SEP-2004) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Oct 5, 2001 this sequence version replaced gi:14670206.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@wustl.edu and jes@sanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

For a graphical representation of this clone sequence and its
analysis see:
<http://www.wormbase.org/db/seq/sequence?name=ZK1236;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is ZK353, 200 bp overlap; the 3' clone is C30C11, 1200
bp overlap. Actual start of this clone is at base position 1 of
ZK1236.

NOTES:

Coding sequences below are the result of integration and manual
review of the following data: computer analysis using the program
GeneFinder (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yuji Kohara
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
elegans ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>),
similarity to other proteins from BlastX analyses
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual C. elegans GenBank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
source

Location/Qualifiers
1..30789
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="ZK1236"
1744..2838
/gene="ZK1236.4"
/locus tag="ZK1236.4"
1744..2838
/gene="ZK1236.4"
/locus tag="ZK1236.4"
/standard name="ZK1236.4"
/note="contains similarity to Pfam domain PF00078 (Reverse
transcriptase (RNA-dependent DNA polymerase))"
/codon_start=1
/product="Hypothetical protein ZK1236.4"
/protein_id="AAA28194.1"
/db_xref="GI:289755"
/db_xref="WormBase:ZK1236.4"
/translation="MLECMNNWTKNIDEGAQTDVKYLDPSKAFDKVSHDILLDKLTSI

gene

CDS

/product="C.elegans chromodomain protein 1"
/protein_id="AAA28192.1"
/db_xref="GI:289753"
/db_xref="WormBase:ZK1236.2"
/translation="MADGSBLYTVESILEHRKKKGKSEFYIKWIGYDHTNSWBPKEN
IVDPTLTFPTFAAKAEIKAKKQWAAKKGKASASASASTPARGAKAA
PKPPKPKSPKQRQIRAGSDIRPDSDTDEHSSADKSKAEDEEVEDDEEPVPPKKKE

Query Match 100.0%; Score 14; DB 3; Length 30789;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TTGCCCCGCCCTT 14

|||||

KINKHLIRWLDFLTNRSFKVGVNTLSEPKTKVCGVPGQSVISPVLFGI FVNEISAN
LPVGYCKQFADDIKLAAAI PKNOSQISLOAIDVVTDWALNNKLSPSKTFLHLTG
KNTNNTYFENGLPIKEVLTFRDGLFVSEKLDFTDHWKKTINTIKYOLFOLIENTYSS
KSVKMLTYLTYTFFIRPLEYGTVTISPIKKNDSKAIESVQNAFTRRFYSRMKGRI RP
DDKDYKASERNELFKLUSTSENRLAIDKLIKLSHMMTKIGLGTSEFPNFTENTRTS
KYKFSWCKCKTITRRFVINTLTSLQNNF"
3336..9533
/gene="ZK1236.3"
/locus tag="ZK1236.3"
join(3336..3503,3828..3916,4055..4364,6314..6474,
6688..7966,8022..8438,8902..9429,9483..9533)
/gene="ZK1236.3"
/locus tag="ZK1236.3"
/standard name="ZK1236.3"
/note="contains similarity to Ashbya gossypii ACR285Cp.;
TR:Q75BI6; coded for by the following C. elegans cDNAs:
CK581370, YK296d8.3, YK296d8.5, YK336g5.3, YK336g5.5,
YK431d5.5, YK526e9.5, YK1279g05.3, YK1279g05.5,
YK1572a12.3, YK1572a12.5, YK1331c08.3, YK1331c08.5"
/codon_start=1
/product="Hypothetical protein ZK1236.3"
/protein_id="AAA28193.1"
/db_xref="GI:289754"
/db_xref="WormBase:ZK1236.3"
/translation="MTINIKYSKFSSTSSSEELPKTYIPAYQPPVGMPPKYVYN
WLRTLSLKNLKNIRAIYLPDQCONFYQESSRKFSGAGOVSSLTPTNYSNFSYTVI
LRWKDTSLLPLENSNGTYLFLNLTVPFSPALSFNYSWRGDTGPSISIFSLSVLV
FSLFPQHKNICARWCPFRRLSLSFSLRIMSSEPASSTSEKVPFEEPHPSIKHKFG
PQVIFRALSDYLVNANNQTPSEYERKALNAKYGRDDYITLCLHVTSLCTSEGLDDGP
EYVLLCDSRARDKLIIDLVKLEIDTDYVQLELHGKGRHQVQVYVHQHOAMNDQSSS
EGSDFTLEMKLVPSSEKAKIMKQEBEERKARKHGOVQOYVHQHOAMNDQSSS
SVPTSSPSSCSSEANREKMETVREPAGPSRLMRAINAPVAPVVIETPVALPEED
ETLMDDEMPSLTVEASEASFEAQSPSPQVPOASIEGSPSQQOI PGTSSQKQVAR
GSRNTMISYHDLPGTGNAPPMACPVTLKLEKNVPFEAKIRVAGIYTRKFI SEVQKM
RPSDLESIFHSICIASVQRIRKRNELVQQLQEI NAQCSFTMTNKKTKLAKAYQRV
QNEIEKIDREGQILPQOYNNMPPMPQGGQRLPPPAYPGILPQOQRNQQGVPPQFQRS
POFMIGDQRYAHYPMQLPNSNORALINTSVQPSVEVNRNLVKTEAMAMNAQLN
PPRPPPPPHRALQGELOFLRPGAPDPCPRDPSKOTYNTYTVTASPATLTNSIIP
WHPPYKSGRLNVSNITKALNEYRLLCNLRQADPASFLPEYFLGDDMPHFNKILSIA
DYNLTSKRCKDEADYIHRMSHSDQLYLLELQSDSNVEKWKTFYRINQWDLPLN
TNEATPRTSSPASNRPTTSTAQPPTLNPT"
13447..15372
/gene="cec-1"
/locus tag="ZK1236.2"
join(13447..13506,13727..13770,13944..14577,15196..15372)
/gene="cec-1"
/locus tag="ZK1236.2"
/standard name="ZK1236.2"
/note="contains similarity to Pfam domain PF00385
('chromo' (Chromatin Organisation Modifier) domain);
coded for by the following C. elegans cDNAs: CESM73,
OSTR056E11.1, OSTR056E11.1, YK104c2.3, YK104c2.5,
YK108h5.3, YK108h5.5, YK116d10.3, YK116d10.5, YK156g10.3,
YK156g10.5, YK172h2.3, YK186f10.5, YK293g5.5, YK371c1.3,
YK371c1.5, YK417a5.3, YK417a5.5, YK427c6.3, YK466f1.3,
YK477c9.3, YK565f2.3, YK565f2.5, YK593g4.3, YK598h11.3,
YK604c1.3, YK606d3.3, YK640b12.3, YK1745c02.5, CESM73f"
/codon_start=1
/product="C.elegans chromodomain protein 1"
/protein_id="AAA28192.1"
/db_xref="GI:289753"
/db_xref="WormBase:ZK1236.2"

Db 16117 TTGCCCCGCCCTT 16130

Search completed: April 25, 2005, 12:19:07
Job time : 321.448 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 11:07:14 ; Search time 633.103 Seconds
(without alignments)
2296.080 Million cell updates/sec

Title: US-10-010-476-10
Perfect score: 30
Sequence: 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	30	100.0	30	6	BD0781125
C 2	30	100.0	30	6	BD0781126
C 3	20	100.0	50	6	BD0781129
C 4	22.6	75.3	201933	2	AC141729
C 5	22	73.3	22	6	BD0781130
C 6	21.6	72.0	705	6	AX5255599
C 7	21.6	72.0	4562	10	D88672
C 8	21.6	72.0	237359	2	AC127753
C 9	21.6	72.0	241243	2	AC094206
C 10	21.6	72.0	263882	2	AC107344
C 11	21.2	70.7	92969	2	AC010844
C 12	21.2	70.7	206000	10	AC102248
C 13	21	70.0	161580	2	AC141660
C 14	21	70.0	208558	2	AC128403
C 15	21	70.0	248338	2	AC106101
C 16	20.8	69.3	248825	2	AC131555
C 17	20.8	69.3	255345	2	AC114514
C 18	20.6	68.7	2209	10	BC021464
C 19	20.6	68.7	2210	10	BC021392

c 93	19.6	65.3	170945	2	AC079911	Mus muscu	AC079911	19.2	64.0	162415	2	AC010065	AC010065 Drosophill
c 94	19.6	65.3	181904	3	AC008211	Drosophill	AC008211	19.2	64.0	162415	2	AC010570	AC010570 Drosophill
c 95	19.6	65.3	210636	9	AC006443	Homo sapi	AL772366	19.2	64.0	167389	10	AL772366	Mouse DNA
c 96	19.6	65.3	213688	3	AE003755	Drosophill	AE003755	19.2	64.0	167630	3	AC009373	AC009373 Drosophill
c 97	19.6	65.3	220838	10	AL645723	Mouse DNA	AL645723	19.2	64.0	174118	3	AC006091	AC006091 Drosophill
c 98	19.6	65.3	235877	2	AC117278	Rattus no	AC117278	19.2	64.0	175906	10	AC125126	AC125126 Mus muscu
c 99	19.6	65.3	272265	33	AC137042	Rattus no	AC137042	19.2	64.0	181587	2	AC150929	AC150929 Bos tauru
c 100	19.4	64.7	33	6	AR139884	Sequence	AR139884	19.2	64.0	183982	2	AC121700	AC121700 Rattus no
c 101	19.4	64.7	33	6	AR167528	Sequence	AR167528	19.2	64.0	190186	4	AC150544	AC150544 Bos tauru
c 102	19.4	64.7	33	6	AR234352	Sequence	AR234352	19.2	64.0	197751	2	AC140075	AC140075 Rattus no
c 103	19.4	64.7	33	6	AR476169	Sequence	AR476169	19.2	64.0	200509	2	AC006910	AC006910 Caenorhab
c 104	19.4	64.7	33	6	AR488054	Sequence	AR488054	19.2	64.0	204448	2	AC123397	AC123397 Rattus no
c 105	19.4	64.7	33	6	BD084556	Recombina	BD084556	19.2	64.0	217092	2	AL159068	AL159068 Mouse DNA
c 106	19.4	64.7	1041	6	AX820522	Sequence	AX820522	19.2	64.0	218889	10	AC079959	AC079959 Mus muscu
c 107	19.4	64.7	1041	6	AX820522	Sequence	AX820522	19.2	64.0	226054	2	AC117359	AC117359 Rattus no
c 108	19.4	64.7	1041	6	AX831552	Sequence	AX831552	19.2	64.0	226360	2	AC138353	AC138353 Mus muscu
c 109	19.4	64.7	1041	8	AY692911	Saccharomyc	AY692911	19.2	64.0	248822	2	AC137460	AC137460 Rattus no
c 110	19.4	64.7	1720	8	SCU85960	BC066993	BC066993	19.2	64.0	249364	2	AC098392	AC098392 Rattus no
c 111	19.4	64.7	3172	10	SC9571X	249810 S.cerevisia	249810	19.2	64.0	251271	2	AC129644	AC129644 Rattus no
c 112	19.4	64.7	29366	8	SC9571X	AF177767	AF177767	19.2	64.0	253323	3	AE003725	AE003725 Drosophill
c 113	19.4	64.7	60810	9	AC025227	Homo sapi	AC025227	19.2	64.0	253942	2	AC096037	AC096037 Rattus no
c 114	19.4	64.7	62899	2	AC134779	Homo sapi	AC134779	19.2	64.0	255062	10	AC122912	AC122912 Mus muscu
c 115	19.4	64.7	72319	2	AC139642	Rattus no	AC139642	19.2	64.0	256306	2	AC117159	AC117159 Rattus no
c 116	19.4	64.7	76734	2	AC023218	Homo sapi	AC023218	19.2	64.0	257486	2	AC125854	AC125854 Rattus no
c 117	19.4	64.7	86621	9	AC003675	Human Chr	AC003675	19.2	64.0	263939	2	AC095238	AC095238 Rattus no
c 118	19.4	64.7	91748	10	AF177767	Mus muscu	AF177767	19.2	64.0	263961	2	AC094811	AC094811 Rattus no
c 119	19.4	64.7	96593	6	AX695776	Sequence	AX695776	19.2	64.0	266687	2	AC118122	AC118122 Rattus no
c 120	19.4	64.7	104123	9	HSB6345	Continuation (4 of	Continuation (4 of	19.2	64.0	273754	2	AC120615	AC120615 Rattus no
c 121	19.4	64.7	115749	9	AC092995	Homo sapi	AC092995	19.2	64.0	328676	3	AE003524	AE003524 Drosophill
c 122	19.4	64.7	118722	2	AP000584	AP000584	AP000584	19.2	64.0	328676	3	AE003524	AE003524 RPNMMSQ00
c 123	19.4	64.7	137932	9	AC005950	Homo sapi	AC005950	19.2	64.0	383	11	BV096381	BV096381 RPNMMSQ00
c 124	19.4	64.7	141033	2	AC143253	Macaca mu	AC143253	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 125	19.4	64.7	144257	2	AC068131	Homo sapi	AC068131	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 126	19.4	64.7	147109	10	AC139183	Mus muscu	AC139183	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 127	19.4	64.7	148894	8	AC121489	AC121489	AC121489	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 128	19.4	64.7	158431	10	AC137901	AC137901	AC137901	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 129	19.4	64.7	158431	2	AC147528	AC147528	AC147528	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 130	19.4	64.7	164628	10	AC124787	AC124787	AC124787	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 131	19.4	64.7	165264	2	AC142254	Mus muscu	AC142254	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 132	19.4	64.7	165333	9	BSX46492	Human DNA	BSX46492	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 133	19.4	64.7	167754	8	AP004303	AP004303	AP004303	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 134	19.4	64.7	171419	9	AC0135142	Rattus no	AC0135142	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 135	19.4	64.7	174484	9	AP002815	Homo sapi	AP002815	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 136	19.4	64.7	174569	9	AC013791	Homo sapi	AC013791	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 137	19.4	64.7	177110	2	AC133042	Homo sapi	AC133042	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 138	19.4	64.7	177302	10	AC127415	Mus muscu	AC127415	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 139	19.4	64.7	181214	2	AC110907	AC110907	AC110907	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 140	19.4	64.7	181904	2	AC084422	Homo sapi	AC084422	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 141	19.4	64.7	182643	9	CNS010DT7	AL132642	AL132642	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 142	19.4	64.7	185235	2	AC012661	Homo sapi	AC012661	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 143	19.4	64.7	186773	9	AC108724	AC108724	AC108724	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 144	19.4	64.7	186861	9	AL592170	AL592170	AL592170	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 145	19.4	64.7	189892	2	AC091425	AC091425	AC091425	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 146	19.4	64.7	191144	9	AC138773	AC138773	AC138773	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 147	19.4	64.7	194627	10	AC150685	AC150685	AC150685	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 148	19.4	64.7	198189	2	AC145043	Papio anu	AC145043	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 149	19.4	64.7	201124	2	AC150228	AC150228	AC150228	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 150	19.4	64.7	209828	10	AC127594	AC127594	AC127594	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 151	19.4	64.7	210892	2	AC102611	AC102611	AC102611	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 152	19.4	64.7	212636	10	AL646097	Mouse DNA	AL646097	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 153	19.4	64.7	214260	2	AC016464	Mus muscu	AC016464	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 154	19.4	64.7	216851	2	AC118154	Rattus no	AC118154	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 155	19.4	64.7	227538	10	AC002397	AC002397	AC002397	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 156	19.4	64.7	231407	2	AC139592	AC139592	AC139592	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 157	19.4	64.7	242454	9	HSAC001228	AC001228	AC001228	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 158	19.4	64.7	251021	2	AC129763	Rattus no	AC129763	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 159	19.4	64.7	300600	1	AP005952	Bradyrhiz	AP005952	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 160	19.4	64.7	327752	2	AC120427	Mus muscu	AC120427	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 161	19.2	64.0	576	11	G94468	S208P633RE	G94468	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 162	19.2	64.0	49623	3	AC087232	Caenorhab	AC087232	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 163	19.2	64.0	68723	2	AC019515	Drosophill	AC019515	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 164	19.2	64.0	69260	2	AC017374	Drosophill	AC017374	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00
c 165	19.2	64.0	152788	2	AC120512	Rattus no	AC120512	19.2	64.0	383	11	BV157856	BV157856 RPNMMSQ00

C 239	19	63.3	201664	10	AL731818	Mouse DNA	312	18.8	62.7	137034	9	AL731568	AL731568 Human DNA
C 240	19	63.3	202421	9	AC091038	Homo sapi	C 313	18.8	62.7	141675	2	AP000405	AP000405 Homo sapi
C 241	19	63.3	223097	2	AC131769	Rattus no	C 314	18.8	62.7	142549	2	AP004271	AP004271 Oryza sat
C 242	19	63.3	223863	2	AC136532	Rattus no	C 315	18.8	62.7	143291	9	HS16339	AL008733 Human DNA
C 243	19	63.3	227781	2	AC132915	Mus muscu	C 316	18.8	62.7	143291	9	HS16339	AL008733 Human DNA
C 244	19	63.3	232763	2	AC097686	Rattus no	C 317	18.8	62.7	144872	9	AC060813	AC060813 Homo sapi
C 245	19	63.3	240481	3	AE003830	Rattus no	C 318	18.8	62.7	147224	2	AP003779	AP003779 Homo sapi
C 246	19	63.3	241077	2	AC125683	Drosophill	C 319	18.8	62.7	148257	10	AL669932	AL669932 Mouse DNA
C 247	19	63.3	250073	2	AC125683	Rattus no	C 320	18.8	62.7	150162	9	AC026722	AC026722 Homo sapi
C 248	19	63.3	250870	2	AC098913	Rattus no	C 321	18.8	62.7	151037	9	AC092620	AC092620 Homo sapi
C 249	19	63.3	266396	2	AC112584	Rattus no	C 322	18.8	62.7	151370	2	CR626879	CR626879 Danio rer
C 250	19	63.3	273225	2	AC108524	Rattus no	C 323	18.8	62.7	152177	10	AL672064	AL672064 Mouse DNA
C 251	19	63.3	274799	2	AC109721	Rattus no	C 324	18.8	62.7	153533	2	AC118972	AC118972 Rattus no
C 252	19	63.3	275390	3	AE003831	Drosophill	C 325	18.8	62.7	153730	2	AL158162	AL158162 Homo sapi
C 253	19	63.3	279252	2	AC093941	Rattus no	C 326	18.8	62.7	157122	9	AC079015	AC079015 Homo sapi
C 254	19	63.3	295650	1	BX294151	Pirellulla	C 327	18.8	62.7	157929	10	AC124407	AC124407 Mus muscu
C 255	19	63.3	295650	1	BX294151	Pirellulla	C 328	18.8	62.7	158989	2	AC068853	AC068853 Homo sapi
C 256	19	63.3	299015	2	AC006842	Caenorhab	C 329	18.8	62.7	159100	2	AC073224	AC073224 Homo sapi
C 257	19	63.3	299782	2	AC006844	Caenorhab	C 330	18.8	62.7	160739	9	AP001092	AP001092 Homo sapi
C 258	18.8	62.7	30	6	BD078127	Modulator	C 331	18.8	62.7	163604	9	AP006301	AP006301 Homo sapi
C 259	18.8	62.7	30	6	BD078128	Modulator	C 332	18.8	62.7	163797	2	AC025872	AC025872 Homo sapi
C 260	18.8	62.7	353	6	CQ748594	Sequence	C 333	18.8	62.7	164735	2	AC090475	AC090475 Homo sapi
C 261	18.8	62.7	656	6	BD149046	Primer fo	C 334	18.8	62.7	164664	9	AL158212	AL158212 Human DNA
C 262	18.8	62.7	656	6	BD149046	Primer fo	C 335	18.8	62.7	165979	9	AC061708	AC061708 Homo sapi
C 263	18.8	62.7	882	14	RWJ225231	Sequence	C 336	18.8	62.7	167284	9	AC112719	AC112719 Homo sapi
C 264	18.8	62.7	1308	9	AF522996	Homo sapi	C 337	18.8	62.7	167854	2	AC092483	AC092483 Homo sapi
C 265	18.8	62.7	1319	9	BC018415	Homo sapi	C 338	18.8	62.7	168991	2	AC025462	AC025462 Homo sapi
C 266	18.8	62.7	1617	6	BD156711	Primer fo	C 339	18.8	62.7	169259	9	AP006288	AP006288 Homo sapi
C 267	18.8	62.7	1617	6	AX877541	Sequence	C 340	18.8	62.7	169695	9	AC091836	AC091836 Homo sapi
C 268	18.8	62.7	1617	9	AX001787	Homo sapi	C 341	18.8	62.7	170114	9	AC044790	AC044790 Homo sapi
C 269	18.8	62.7	1980	9	BC002749	Homo sapi	C 342	18.8	62.7	172260	9	AC068946	AC068946 Homo sapi
C 270	18.8	62.7	2021	6	AX746924	Sequence	C 343	18.8	62.7	172315	9	AC018553	AC018553 Homo sapi
C 271	18.8	62.7	2021	6	AX091345	Homo sapi	C 344	18.8	62.7	173661	2	AC074107	AC074107 Mus muscu
C 272	18.8	62.7	2040	9	HS4417059	Homo sapi	C 345	18.8	62.7	173696	10	AL670603	AL670603 Mouse DNA
C 273	18.8	62.7	2048	6	AX713453	Sequence	C 346	18.8	62.7	173837	10	AC124681	AC124681 Mus muscu
C 274	18.8	62.7	2048	8	AK054872	Homo sapi	C 347	18.8	62.7	174458	9	AC091189	AC091189 Homo sapi
C 275	18.8	62.7	2248	8	AK054872	Homo sapi	C 348	18.8	62.7	175100	9	AC092818	AC092818 Homo sapi
C 276	18.8	62.7	2563	9	BC015722	Homo sapi	C 349	18.8	62.7	176118	2	AC068670	AC068670 Homo sapi
C 277	18.8	62.7	2573	1	AB003380	Fibroblast	C 350	18.8	62.7	176584	2	AC062010	AC062010 Homo sapi
C 278	18.8	62.7	2875	5	BC064723	Xenopus t	C 351	18.8	62.7	176814	9	AC100849	AC100849 Homo sapi
C 279	18.8	62.7	3150	1	AF013987	Vibrio ch	C 352	18.8	62.7	177076	9	AP000487	AP000487 Homo sapi
C 280	18.8	62.7	11328	1	AE004322	Vibrio ch	C 353	18.8	62.7	177958	9	AC114933	AC114933 Homo sapi
C 281	18.8	62.7	38902	9	AC010643	Homo sapi	C 354	18.8	62.7	178245	2	AC143323	AC143323 Homo sapi
C 282	18.8	62.7	46758	2	AC151843	Continuation (6 of	C 355	18.8	62.7	178804	2	AC108418	AC108418 Mus muscu
C 283	18.8	62.7	50341	6	AR007269	Sequence	C 356	18.8	62.7	179075	2	AC150138	AC150138 Gallus ga
C 284	18.8	62.7	50341	6	AR007269	Sequence	C 357	18.8	62.7	179883	2	AC007721	AC007721 Homo sapi
C 285	18.8	62.7	52297	6	AR148205	Sequence	C 358	18.8	62.7	183402	2	AC126231	AC126231 Bos tauru
C 286	18.8	62.7	52297	6	AR369021	Sequence	C 359	18.8	62.7	184092	2	AC124858	AC124858 Homo sapi
C 287	18.8	62.7	52297	7	MLCGA	Sequence	C 360	18.8	62.7	184143	2	AC141993	AC141993 Rattus no
C 288	18.8	62.7	62889	2	AC099938	Mus muscu	C 361	18.8	62.7	184237	2	AC108418	AC108418 Mus muscu
C 289	18.8	62.7	64501	2	AC068149	Homo sapi	C 362	18.8	62.7	184708	2	AC141454	AC141454 Homo sapi
C 290	18.8	62.7	70172	2	AC130370	Homo sapi	C 363	18.8	62.7	185588	9	AP006305	AP006305 Homo sapi
C 291	18.8	62.7	87256	9	AC091865	Homo sapi	C 364	18.8	62.7	185860	9	AC026471	AC026471 Homo sapi
C 292	18.8	62.7	94045	2	AC010039	Drosophill	C 365	18.8	62.7	188549	9	CNS05TE4	AL358333 Human chr
C 293	18.8	62.7	94212	2	AC091848	Homo sapi	C 366	18.8	62.7	188833	9	HS268TE5	AL358333 Human chr
C 294	18.8	62.7	99774	10	AL772379	Mouse DNA	C 367	18.8	62.7	188896	9	AL136221	AL136221 Human DNA
C 295	18.8	62.7	100000	9	AP000072	Homo sapi	C 368	18.8	62.7	189381	2	AC126920	AC126920 Bos tauru
C 296	18.8	62.7	100000	2	AC139486	Homo sapi	C 369	18.8	62.7	189810	9	AC068880	AC068880 Homo sapi
C 297	18.8	62.7	110000	2	AC151828	Continuation (3 of	C 370	18.8	62.7	191334	2	AC093385	AC093385 Homo sapi
C 298	18.8	62.7	110000	2	AC151846	Continuation (3 of	C 371	18.8	62.7	191898	9	AP003071	AP003071 Homo sapi
C 299	18.8	62.7	112144	8	AP003758	Oryza sat	C 372	18.8	62.7	192239	9	AC009796	AC009796 Homo sapi
C 300	18.8	62.7	113170	8	AP003804	Oryza sat	C 373	18.8	62.7	192365	2	AC125029	AC125029 Mus muscu
C 301	18.8	62.7	114089	10	AL807770	Mouse DNA	C 374	18.8	62.7	194254	10	AC127312	AC127312 Mus muscu
C 302	18.8	62.7	116986	2	AC150107	Gallus ga	C 375	18.8	62.7	194781	2	AC019124	AC019124 Homo sapi
C 303	18.8	62.7	118631	9	AC104691	Homo sapi	C 376	18.8	62.7	195766	9	CNS01DVC	AL135744 Human chr
C 304	18.8	62.7	119118	9	HSRY7CC1	Homo sapi	C 377	18.8	62.7	199241	2	AC146804	AC146804 Bos tauru
C 305	18.8	62.7	119118	9	HSRY7CC1	Homo sapi	C 378	18.8	62.7	199258	10	AC127356	AC127356 Mus muscu
C 306	18.8	62.7	119151	2	AP000655	Homo sapi	C 379	18.8	62.7	199386	9	AC117440	AC117440 Homo sapi
C 307	18.8	62.7	120628	9	AC012153	Homo sapi	C 380	18.8	62.7	199595	2	AC118936	AC118936 Homo sapi
C 308	18.8	62.7	120977	5	BS39334	Zebrafish	C 381	18.8	62.7	201294	2	AC128748	AC128748 Rattus no
C 309	18.8	62.7	131152	10	AL929068	Mouse DNA	C 382	18.8	62.7	202356	2	CR354373	CR354373 Danio rer
C 310	18.8	62.7	132764	9	AC073842	Homo sapi	C 383	18.8	62.7	203897	10	AC132434	AC132434 Mus muscu
C 311	18.8	62.7	135991	2	AC148750	Dasyypus n	C 384	18.8	62.7	204540	9	AC122710	AC122710 Homo sapi

385	18.8	62.7	205066	2	AC120803	AC120803 Rattus no	C 458	18.6	62.0	2410	9	BC036849	BC036849 Homo sapi
386	18.8	62.7	205476	2	AC136731	AC136731 Mus muscu	459	18.6	62.0	3230	5	BC065959	BC065959 Danio rer
387	18.8	62.7	207418	10	AC131104	AC131104 Mus muscu	C 460	18.6	62.0	4375	9	AK090460	AK090460 Homo sapi
388	18.8	62.7	208430	2	AC027641	AC027641 Homo sapi	C 461	18.6	62.0	21925	6	AX647863	AX647863 Sequence
389	18.8	62.7	209586	2	AC148497	AC148497 Otolomur	C 462	18.6	62.0	26184	6	AX647497	AX647497 Sequence
C 390	18.8	62.7	209736	10	AC117242	AC117242 Mus muscu	C 463	18.6	62.0	37644	6	A95302	A95302 Sequence 33
C 391	18.8	62.7	210202	2	AC008778	AC008778 Homo sapi	464	18.6	62.0	66772	2	AC123762	AC123762 Mus muscu
C 392	18.8	62.7	210402	9	AP001271	AP001271 Homo sapi	C 465	18.6	62.0	68939	10	BX544888	BX544888 Mouse DNA
C 393	18.8	62.7	210509	10	AC105073	AC105073 Mus muscu	C 466	18.6	62.0	69004	9	AL355145	AL355145 Human DNA
C 394	18.8	62.7	211117	2	AC132627	AC132627 Rattus no	467	18.6	62.0	79175	9	AC025768	AC025768 Homo sapi
C 395	18.8	62.7	211132	2	AC145337	AC145337 Pan trogl	C 468	18.6	62.0	92863	12	GBLOCUS	AL023516 Gallus ga
C 396	18.8	62.7	211132	2	AC145337	AC145337 Pan trogl	469	18.6	62.0	95599	6	AX695563	AX695563 Sequence
C 397	18.8	62.7	211735	9	AC091564	AC091564 Homo sapi	470	18.6	62.0	110000	2	AC102394	Continuaton (2 of
C 398	18.8	62.7	211741	2	AC102907	AC102907 Mus muscu	471	18.6	62.0	114156	10	AL590390	AL590390 Mouse DNA
C 399	18.8	62.7	213948	2	AC139346	AC139346 Mus muscu	C 472	18.6	62.0	132419	1	D90907	D90907 Synectocyt
C 400	18.8	62.7	213925	2	AC150641	AC150641 Bos tauru	473	18.6	62.0	135554	2	AC116945	AC116945 Tetraodon
401	18.8	62.7	214182	2	AC113634	AC113634 Rattus no	474	18.6	62.0	137284	2	AC148774	AC148774 Mus muscu
C 402	18.8	62.7	215505	10	AL844166	AL844166 Mouse DNA	C 475	18.6	62.0	138435	9	AC073468	AC073468 Homo sapi
C 403	18.8	62.7	215671	10	AC113018	AC113018 Mus muscu	476	18.6	62.0	140001	9	AC090012	AC090012 Homo sapi
C 404	18.8	62.7	216935	2	AC132648	AC132648 Rattus no	477	18.6	62.0	154235	9	AC123029	AC123029 Mus muscu
C 405	18.8	62.7	216980	2	AC106640	AC106640 Rattus no	C 478	18.6	62.0	155606	9	AC091851	AC091851 Homo sapi
C 406	18.8	62.7	217714	2	CR762397	CR762397 Danio rer	C 479	18.6	62.0	160311	9	AL354889	AL354889 Human DNA
407	18.8	62.7	218067	2	AC150058	AC150058 Gallus ga	C 480	18.6	62.0	162445	9	AL158151	AL158151 Human DNA
C 408	18.8	62.7	221452	9	AC034242	AC034242 Homo sapi	C 481	18.6	62.0	163283	2	AC112980	AC112980 Mus muscu
C 409	18.8	62.7	223183	2	AC131405	AC131405 Rattus no	482	18.6	62.0	164883	10	AL831756	AL831756 Mouse DNA
C 410	18.8	62.7	223952	10	AC107742	AC107742 Mus muscu	C 483	18.6	62.0	167480	9	H8BA18114	AL121928 Human DNA
C 411	18.8	62.7	226589	2	AC117299	AC117299 Rattus no	C 484	18.6	62.0	167727	10	AC101093	AC101093 Mus muscu
C 412	18.8	62.7	227273	2	AC141526	AC141526 Rattus no	485	18.6	62.0	167974	9	AL591073	AL591073 Human DNA
C 413	18.8	62.7	228101	2	AC107406	AC107406 Rattus no	486	18.6	62.0	170011	2	AC1018583	AC1018583 Homo sapi
C 414	18.8	62.7	230637	2	AC095345	AC095345 Rattus no	487	18.6	62.0	170607	2	AC102151	AC102151 Mus muscu
C 415	18.8	62.7	230951	2	AC129271	AC129271 Rattus no	C 488	18.6	62.0	170819	2	AC011780	AC011780 Homo sapi
C 416	18.8	62.7	232958	2	AC135146	AC135146 Rattus no	489	18.6	62.0	171297	2	AC137743	AC137743 Mus muscu
C 417	18.8	62.7	233609	2	AC118453	AC118453 Rattus no	490	18.6	62.0	173522	2	AC073364	AC073364 Homo sapi
C 418	18.8	62.7	237878	2	AC106542	AC106542 Rattus no	491	18.6	62.0	173672	2	AC121520	AC121520 Mus muscu
C 419	18.8	62.7	240138	2	AC111703	AC111703 Rattus no	C 492	18.6	62.0	174969	2	AC151745	AC151745 Mus muscu
C 420	18.8	62.7	241982	2	AC105822	AC105822 Rattus no	493	18.6	62.0	175262	9	AC021491	AC021491 Homo sapi
C 421	18.8	62.7	242721	2	AC097403	AC097403 Rattus no	C 494	18.6	62.0	175878	2	AC122554	AC122554 Mus muscu
C 422	18.8	62.7	243369	10	AL660973	AL660973 Mouse DNA	495	18.6	62.0	178377	2	CR384053	CR384053 Danio rer
C 423	18.8	62.7	243654	10	AL663082	AL663082 Mouse DNA	496	18.6	62.0	178879	2	AC124052	AC124052 Mus muscu
C 424	18.8	62.7	244171	2	AC105706	AC105706 Rattus no	497	18.6	62.0	183468	10	AC141469	AC141469 Mus muscu
C 425	18.8	62.7	247268	2	AC125571	AC125571 Rattus no	C 498	18.6	62.0	183468	10	AC141469	AC141469 Mus muscu
C 426	18.8	62.7	248308	2	AC123259	AC123259 Rattus no	499	18.6	62.0	185395	2	AC118699	AC118699 Mus muscu
427	18.8	62.7	248672	2	AC129613	AC129613 Rattus no	500	18.6	62.0	186184	2	AC139894	AC139894 Rattus no
C 428	18.8	62.7	252248	2	AC023238	AC023238 Homo sapi							
C 429	18.8	62.7	253523	2	AC126062	AC126062 Rattus no							
C 430	18.8	62.7	253523	2	AC096066	AC096066 Rattus no							
431	18.8	62.7	261511	2	AC098318	AC098318 Rattus no							
C 432	18.8	62.7	261686	2	AC124315	AC124315 Mus muscu							
C 433	18.8	62.7	263879	2	AC096060	AC096060 Rattus no							
C 434	18.8	62.7	264814	2	AC122631	AC122631 Rattus no							
C 435	18.8	62.7	267398	2	AC106519	AC106519 Rattus no							
C 436	18.8	62.7	271599	2	AC095812	AC095812 Rattus no							
437	18.8	62.7	274343	2	AC115208	AC115208 Rattus no							
C 438	18.8	62.7	280575	2	AC117364	AC117364 Rattus no							
C 439	18.8	62.7	300217	1	AE016922	AE016922 Chromobac							
C 440	18.8	62.7	304407	2	AC073813	AC073813 Mus muscu							
C 441	18.8	62.7	313378	2	AC094200	AC094200 Rattus no							
C 442	18.8	62.7	315370	2	AC109390	AC109390 Rattus no							
C 443	18.8	62.7	345630	2	AC098229	AC098229 Rattus no							
444	18.8	62.7	346027	2	AC128985	AC128985 Rattus no							
445	18.6	62.0	180	6	CQ662027	CQ662027 Sequence							
C 446	18.6	62.0	240	9	HS97G9R	Z64138 H. sapiens C							
C 447	18.6	62.0	262	6	CQ697578	CQ697578 Sequence							
448	18.6	62.0	277	6	CQ664542	CQ664542 Sequence							
449	18.6	62.0	425	14	AV131352	AV131352 Hepatitis							
450	18.6	62.0	425	14	AV131353	AV131353 Hepatitis							
451	18.6	62.0	425	14	AV131358	AV131358 Hepatitis							
452	18.6	62.0	528	11	BV209823	BV209823 CHAD 4562							
453	18.6	62.0	940	9	HSCHON03	U96769 Homo sapien							
C 454	18.6	62.0	1428	1	AF306346	AF306346 Mycoplasma							
455	18.6	62.0	1708	9	BC036360	BC036360 Homo sapi							
456	18.6	62.0	1722	4	BOVLDLRX	M1341 Bovine low							
457	18.6	62.0	1739	9	BC073974	BC073974 Homo sapi							

ALIGNMENTS

RESULT 1	BD078125/c	BD078125	30 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	Modulator of DNA cytosine-5 methyltransferase and method of using	the same.				
DEFINITION	the same.					
ACCESSION	BD078125	GI:22623728				
VERSION	JP 2001514862-A/9.					
KEYWORDS	unidentified					
SOURCE	unclassified.					
ORGANISM	1 (bases 1 to 30)					
REFERENCE	Reich,N.O. and FLYNN,J.					
AUTHORS	Modulator of DNA cytosine-5 methyltransferase and method of using					
TITLE	the same					
JOURNAL	Patent: JP 2001514862-A 9 18-SEP-2001-					
COMMENT	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA					
	OS Unidentified					
	PN JP 2001514862-A/9					
	PD 18-SEP-2001					
	PF 12-JUN-1998 JP 2000509713					
	PR 29-AUG-1997 US 60/057411					
	PI NORBERT O REICH, JAMES FLYNN					
	PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00					
	CC 'Strandedness: Double;					


```

CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
CC using the same
CC Key Location/Qualifiers
FH source 1..30
FT /organism='Unidentified'.
FT Location/Qualifiers
FEATURES
source
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 30 CTGGATCCTTGGCCCGCCCTTGAATCCC 1

RESULT 2
BD078126 30 bp DNA linear PAT 27-AUG-2002
LOCUS Modulator of DNA cytosine-5 methyltransferase and method of using
DEFINITION the same.
ACCESSION BD078126
VERSION BD078126.1 GI:22623729
KEYWORDS JP 2001514862-A/10.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT PN JP 2001514862-A/10
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000505741
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH, JAMES FLYNN
PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT source 1..30
FT /organism='Unidentified'.
FT Location/Qualifiers
FEATURES
source
1..30
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30

RESULT 3
BD078129 50 bp DNA linear PAT 27-AUG-2002
LOCUS Modulator of DNA cytosine-5 methyltransferase and method of using
DEFINITION the same.
ACCESSION BD078129
VERSION BD078129.1 GI:29123913
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Apis mellifera (honey bee)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apoidea; Apoidea;
Apidae; Apis.
REFERENCE 1 (bases 1 to 201933)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alabrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferracuto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

```

```

the same.
BD078129 1 GI:22623732
KEYWORDS JP 2001514862-A/13.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 50)
AUTHORS Reich,N.O. and Flynn,J.
TITLES Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL Patent: JP 2001514862-A 13 18-SEP-2001;...
COMMENT OS Unidentified
PN JP 2001514862-A/13
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000505741
PR 29-AUG-1997 US 60/057411
PI NORBERT O REICH, JAMES FLYNN
PC C12N9/10, C12Q1/48//C12N15/09, C12N15/00
CC Strandedness: Double;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT source 1..50
FT /organism='Unidentified'.
FT Location/Qualifiers
FEATURES
source
1..50
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGATCCTTGGCCCGCCCTTGAATCCC 30
Db 11 CTGGATCCTTGGCCCGCCCTTGAATCCC 40

RESULT 4
AC141729 201933 bp DNA linear HTG 19-MAR-2003
LOCUS Apis mellifera clone CH224-57G1, WORKING DRAFT SEQUENCE, 44
DEFINITION unordered pieces.
ACCESSION AC141729
VERSION AC141729.1 GI:29123913
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Apis mellifera (honey bee)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apoidea; Apoidea;
Apidae; Apis.
REFERENCE 1 (bases 1 to 201933)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alabrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferracuto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

```

Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mathiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Oragunye, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Orawu, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 201933)
 Worley, K.C.
 Direct Submission
 Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: AMEX
 Center clone name: CH24-57G1
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 188944 bases at least Q40
 Consensus quality: 195465 bases at least Q30
 Consensus quality: 199740 bases at least Q20
 Estimated insert size: 188133; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * As soon as it is available and the accession number will
 * be preserved.
 * 1049: contig of 1049 bp in length
 * 1050 1149: gap of unknown length
 * 1150 2453: contig of 1304 bp in length
 * 2454 2553: gap of unknown length
 * 2554 3986: contig of 1433 bp in length
 * 3987 4086: gap of unknown length
 * 4087 5285: contig of 1199 bp in length
 * 5286 5385: gap of unknown length
 * 5386 6665: contig of 1280 bp in length
 * 6666 6765: gap of unknown length
 * 6766 7932: contig of 1167 bp in length
 * 7933 8032: gap of unknown length
 * 8033 9094: contig of 1062 bp in length
 *
 * 9095 9194: gap of unknown length
 * 9195 11010: contig of 1816 bp in length
 * 11011 12567: contig of 1457 bp in length
 * 12568 12667: gap of unknown length
 * 12668 13675: contig of 1008 bp in length
 * 13676 13775: gap of unknown length
 * 13776 14792: contig of 1017 bp in length
 * 14793 14892: gap of unknown length
 * 14893 16593: contig of 1701 bp in length
 * 16594 18146: contig of 1453 bp in length
 * 16594 18146: contig of 1453 bp in length
 * 18147 18246: gap of unknown length
 * 18247 19468: contig of 1222 bp in length
 * 19469 19568: gap of unknown length
 * 19569 21346: contig of 1778 bp in length
 * 21347 21446: gap of unknown length
 * 21447 22562: contig of 1116 bp in length
 * 22563 22662: gap of unknown length
 * 22663 23992: contig of 1330 bp in length
 * 23993 24092: gap of unknown length
 * 24093 25589: contig of 1497 bp in length
 * 25590 25689: gap of unknown length
 * 25690 27148: contig of 1459 bp in length
 * 27149 27248: gap of unknown length
 * 27249 28226: contig of 1478 bp in length
 * 28227 28826: gap of unknown length
 * 28827 30271: contig of 1445 bp in length
 * 30272 30371: gap of unknown length
 * 30372 31940: contig of 1569 bp in length
 * 31941 32040: gap of unknown length
 * 32041 33598: contig of 1558 bp in length
 * 33599 33698: gap of unknown length
 * 33699 34906: contig of 1208 bp in length
 * 34907 35006: gap of unknown length
 * 35007 36460: contig of 1454 bp in length
 * 36461 36560: gap of unknown length
 * 36561 37821: contig of 1261 bp in length
 * 37822 37921: gap of unknown length
 * 37922 39404: contig of 1483 bp in length
 * 39405 39504: gap of unknown length
 * 39505 42330: contig of 2826 bp in length
 * 42331 42430: gap of unknown length
 * 42431 44150: contig of 1720 bp in length
 * 44151 44250: gap of unknown length
 * 44251 46136: contig of 1886 bp in length
 * 46137 46236: gap of unknown length
 * 46237 49237: contig of 3001 bp in length
 * 49238 49337: gap of unknown length
 * 49338 51357: contig of 2020 bp in length
 * 51358 51457: gap of unknown length
 * 51458 55433: contig of 3976 bp in length
 * 55434 55534: gap of unknown length
 * 55534 59048: contig of 3515 bp in length
 * 59049 59148: gap of unknown length
 * 59149 62792: contig of 3644 bp in length
 * 62793 62892: gap of unknown length
 * 62893 70213: contig of 7321 bp in length
 * 70214 70313: gap of unknown length
 * 70314 78679: contig of 8366 bp in length
 * 78680 89130: contig of 10351 bp in length
 * 89131 89230: gap of unknown length
 * 89231 102519: contig of 13289 bp in length
 * 102520 102619: gap of unknown length
 * 102620 117485: contig of 14866 bp in length
 * 117486 117585: gap of unknown length
 * 117586 132498: contig of 14913 bp in length
 * 132499 132598: gap of unknown length
 * 132599 148663: contig of 16065 bp in length
 * 148664 148763: gap of unknown length
 * 148764 172198: contig of 23435 bp in length
 * 172199 172298: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL

COMMENT

```

* 172299 201933: contig of 29635 bp in length.
FEATURES             source
    Location/Qualifiers
    1..201933
    /organism="Apis mellifera"
    /mol_type="genomic DNA"
    /db_xref="taxon:7460"
    /clone="CH224-57G1"
ORIGIN
Query Match          75.3%; Score 22.6; DB 2; Length 201933;
Best Local Similarity 86.2%; Pred. No. 38;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 29
    |||||
DB 59512 CTGGATCCTTGGCCCGCCCTTGAATTC 59540

RESULT 5
BD078130          22 bp DNA linear PAT 27-AUG-2002
DEFINITION
Modulator of DNA cytosine-5 methyltransferase and method of using
the same.
ACCESSION
BD078130
VERSION
JP 2001514862-A/14.
KEYWORDS
unidentified
SOURCE
unclassified.
ORGANISM
1 (bases 1 to 22)
REFERENCE
Reich,N.O. and Flynn,J.
Modulator of DNA cytosine-5 methyltransferase and method of using
the same
JOURNAL
Patent: JP 2001514862-A 14 18-SEP-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Unidentified
PN JP 2001514862-A/14
PD 18-SEP-2001
PF 12-JUN-1998 JP 2000508978
PI 29-AUG-1997 US 60/057411
PR NORBERT O REICH, JAMES FLYNN
PC C12N9/10.C12Q1/48//C12N15/09.C12N15/00
CC Strandedness: Doublet;
CC Topology: Linear;
CC Modulator of DNA cytosine-5 methyltransferase and method of
using the same
FH Key Location/Qualifiers
FT source 1..22 /organism='Unidentified'.
FEATURES             source
    Location/Qualifiers
    1..22
    /organism="unidentified"
    /mol_type="genomic DNA"
    /db_xref="taxon:32644"
ORIGIN
Query Match          73.3%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ATCCTTGGCCCGCCCTTGAAT 26
    |||||
DB 1 ATCCTTGGCCCGCCCTTGAAT 22

RESULT 6
AX525599          705 bp DNA linear PAT 21-NOV-2002
DEFINITION
Sequence 121 from Patent WO0206682.
ACCESSION
AX525599
VERSION
AX525599.1 GI:25170477
KEYWORDS
Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1
AUTHORS
Farris,G., Hicken,S.H. and Farris,S.B.
TITLE
Rat toxicologically relevant genes and uses thereof
JOURNAL
Patent: WO 0206682-A 121 29-AUG-2002;
Phase-1 Molecular Toxicology Inc. (US)
FEATURES             source
    Location/Qualifiers
    1..705
    /organism="Rattus norvegicus"
    /mol_type="unassigned DNA"
    /db_xref="taxon:10116"
ORIGIN
Query Match          72.0%; Score 21.6; DB 6; Length 705;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 28
    |||||
DB 185 CTGGATCCTTGGCCCGCCCTTGAATTC 212

RESULT 7
D88672          4562 bp mRNA linear ROD 07-FEB-1999
LOCUS
Rattus norvegicus mRNA for phospholipase D, complete cds.
DEFINITION
D88672
ACCESSION
D88672.1 GI:2077942
VERSION
phospholipase D.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (sites)
AUTHORS
Kodaki,T. and Yamashita,S.
TITLE
Cloning, expression, and characterization of a novel phospholipase
D complementary DNA from rat brain
JOURNAL
J. Biol. Chem. 272 (17), 11408-11413 (1997)
MEDLINE
97289050
PUBMED
9111050
REFERENCE
2 (sites)
AUTHORS
Kodaki,T.
TITLE
Cloning and Characterization of rat phospholipase D
JOURNAL
Unpublished
REFERENCE
3 (bases 1 to 4562)
AUTHORS
Kodaki,T.
TITLE
Direct Submission
JOURNAL
Submitted (30-OCT-1996) Tsutomu Kodaki, Kyoto University, Institute
of Advanced Energy; Gokanohs, Uji, Kyoto 611, Japan
(E-mail:kodaki@iae.kyoto-u.ac.jp, Tel:0774-38-3510,
Fax:0774-38-3512)
FEATURES             source
    Location/Qualifiers
    1..4562
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /db_xref="taxon:10116"
    337..3138
    /codon_start=1
    /product="phospholipase D"
    /protein_id="BAA19882.1"
    /db_xref="GI:2077943"
    /translation="MTVTQDLFPYGDVYNSQLHMEPDEVDTLKEGEDPADRMHPFL
    AIYDQELRAHPLVPAPGVPIAQQVGVTERITSGSKVGCTLIYSRLTHGDTWTYTK
    KFRHQELRLDQRHKVLSLPLARFAAHSPAREANENPSLPRGSGSARHTA
    SKQKLYENLRLTFYRNHMTFELVSQSFIPDLGSKGLEGVIRKSGGHRV
    PGFTCCGRDQVCYRWSKRWLVKDSFLLYMRPETGAI SFVQLFDPGFVQVGRSTEA
    RYGVPIDTSRSLILKCSSYRQARWNGQETELAQPGRDFLQHQHDSYAPRPCTL
    ARWVFNAGYFAAVALIRAREEIFITDWLSPILYLRPAHSDDLRLDMLKRAE
    EGRVRSILLFKEVLAIGNSGSKRTLLMLLPNIKVMRHPDLVTLNAHHEKLLVVDQ

```

```

AVAFGLGLDYLAYGRWDDVOYRLTDLGDFSESADSTPTPGSDPAATPDLSHNHPFWLG
KOYSNLYTKWQVLDPRPFDFIDRETTJPMWRDVGWVGVGAARDLARGHTFORWNFT
KTKIARYKIPQVYLPKPSASFANHLPTIIPGAOCATQVQLRSVDRWSAGTLESSILN
AYLHTIRSOHLFYLIENOFFFISCSGRVTLNKVGEIVDRILKAHQOQCRVYVLLP
LLPGFEGDISITGGNSQAILHRTYLCRGSYISHLRLKAAMGAWARDWYSICGLRT
HGLGGEHYISLIYTHSKLLIADDDRTYVIGSANSINDRLLGKSELAILIEDTMEP
SLMDGVEYQAGRFALSRLKHCFSVILGANTWPDLDLDPVCDFFQFLWQETAEANNATI
YEQIFRCLPSNATSLRALRYAVAVESLATVSPSLAQSELAIHRGLHVHFFLPKLEDE
SLLPLGSGKEGMPLEVT"

ORIGIN
Query Match      72.0%; Score 21.6; DB 10; Length 4562;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGATTC 28
DB 3108 CTGGATCTTGGCCACACCCCTGATTC 3215

RESULT 8
AC127753 237359 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-21M19, *** SEQUENCING IN PROGRESS
DEFINITION *** 6 unordered pieces.
AC127753.3 GI:25079334
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
REFERENCE 1 (bases 1 to 237359)
AUTHORS Muzny, D.Marie., Metzker, M.Lee., Abramzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, J., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draeper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Louiseged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelmele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,

```

```

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Tabor, P., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 237359)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237359)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZTT
Center Clone name: CH230-21M19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 186995 bases at least Q40
Consensus quality: 190628 bases at least Q30
Consensus quality: 193100 bases at least Q20
Estimated insert size: 198498; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 51279: contig of 51279 bp in length
* 51280 51379: gap of unknown length
* 51380 153527: contig of 102148 bp in length
* 153528 153627: gap of unknown length
* 153628 179889: contig of 26262 bp in length
* 179890 179989: gap of unknown length
* 179990 212085: contig of 32096 bp in length
* 212086 212185: gap of unknown length
* 212186 235740: contig of 23555 bp in length
* 235741 235840: gap of unknown length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

COMMENT

```

* 235841 237359: Contig of 1519 bp in length.

FEATURES

Location/Qualifiers
1. 237359
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-21M19"
complement(22428..23293)
/note="clone boundary"
note_end:T7
site:
end_sequence:BH272355"

misc_feature

ORIGIN

Query Match 72.0%; Score 21.6; DB 2; Length 237359;

Best Local Similarity 85.7%; Pred. No. 1e+02;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTC 28

DB 157152 CTGGATCTTGGCCCGCCCTGATTC 157179

RESULT 9

AC094206

LOCUS AC094206 241243 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-3A15, WORKING DRAFT SEQUENCE, 4
unorderded pieces.

ACCESSION

AC094206 GI:30467689

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS 1 (bases 1 to 241243)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarone,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshuwa,L., Louisegh,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olampunsaagoo,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H.H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L.,
Puzo,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 241243)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241243)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24819491.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAET

Center clone name: CH230-3A15

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 226414 bases at least Q40

Consensus quality: 229449 bases at least Q30

Consensus quality: 231607 bases at least Q20

Estimated insert size: 233457; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 54907: contig of 54907 bp in length

* 54908 55007: gap of unknown length

* 55008 237646: contig of 182639 bp in length

* 237647 237746: gap of unknown length

* 237747 239762: contig of 2016 bp in length

* 239763 239862: gap of unknown length

* 239863 241243: contig of 1381 bp in length.

FEATURES

Location/Qualifiers
 1. .241243
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-3A15"
 1. .1261
 /note="wgs_contig"
 53814. .54907
 /note="wgs_contig"
 55008. .56017
 /note="wgs_contig"

ORIGIN

Query Match 72.0%; Score 21.6; DB 2; Length 241243;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CTGATCTTCGCGCCGCCCTTGAATTC 28

Db 194119 CTGATCTTCGCGCCGCCCTTCCATTC 194146

RESULT 10

AC107344
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-117F21, linear HTG 22-SEP-2002
 *** 5 unordered pieces.

AC107344

AC107344.4 GI:23195787

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 263882)
 Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsebrook,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gedgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorenshuwa,L., Loulasegh,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwaokemele,O., Okwionu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,D.-L.,
 Puazo,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL
 REFERENCE
 2 (bases 1 to 263882)

Worley,K.C.

TITLE

JOURNAL
 REFERENCE
 Submitted (19-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 263882)

Rat Genome Sequencing Consortium.

TITLE

JOURNAL
 REFERENCE
 Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21737124.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the

sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence

reads. Both end sequences and whole genome shotgun sequence only

contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GOOX

Center clone name: CH230-117F21

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 226341 bases at least Q40

Consensus quality: 229458 bases at least Q30

Consensus quality: 231425 bases at least Q20

Estimated insert size: 248600; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 27027: contig of 27027 bp in length

* 27028 27127: gap of unknown length

* 27128 254735: contig of 227608 bp in length

* 254736 254835: gap of unknown length

* 254836 256430: contig of 1595 bp in length

* 256431 256530: gap of unknown length

* 256531 258485: contig of 1955 bp in length

* 258486 258585: gap of unknown length

* 258586 263882: contig of 5297 bp in length.

Location/Qualifiers

1. .263882

/organism="Rattus norvegicus"

FEATURES

source

```

/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-117F21"
1. .1357
/misc_feature
  /notes="wgs_end_extension"
  clone_end:T7"
  5003_5484
/misc_feature
  /notes="clone_boundary"
  clone_end:T7
  site:EcoRI
  end_sequence:BH280995"
/misc_feature
  27128_31894
  /note="wgs_contig"

ORIGIN
Query Match          72.0%; Score 21.6; DB 2; Length 263882;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAAATTC 28
      |||||
Db 103101 CTGGATCCTTGGCCCGCCACACCTGAGTTC 103128

RESULT 11
AC010844/c
LOCUS
DEFINITION
  AC010844 92969 bp DNA linear HTG 15-FEB-2000
  Drosophila melanogaster chromosome X clone BACR03F03 (D882) RPCI-98
  O3.F.3 map 12C-12D strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
  *** 100 unordered pieces.
ACCESSION
  AC010844.5 GI:6978355
VERSION
  HTG; HTGS PHASE1.
KEYWORDS
  Drosophila melanogaster (fruit fly)
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 92969)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 92969)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 15, 2000 this sequence version replaced gi:6016642.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 100 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 664: contig of 664 bp in length
665 744: gap of unknown length
745 1438: contig of 694 bp in length
1439 1518: gap of unknown length
1519 2157: contig of 639 bp in length
2158 2237: gap of unknown length
2238 2865: contig of 628 bp in length
2866 2946: gap of unknown length
2946 3570: contig of 624 bp in length
3570 3650: gap of unknown length
3650 4397: contig of 748 bp in length
4398 4477: gap of unknown length
4478 5068: contig of 591 bp in length
5069 5148: gap of unknown length
5149 5702: contig of 554 bp in length
5703 5782: gap of unknown length
5783 6679: contig of 896 bp in length
6679 7842: contig of 1084 bp in length
7843 7922: gap of unknown length
7923 8819: contig of 897 bp in length
8820 8899: gap of unknown length
8900 9695: contig of 796 bp in length
9696 9775: gap of unknown length
9776 10180: contig of 405 bp in length
10181 10260: gap of unknown length
10261 10980: contig of 720 bp in length
10981 11060: gap of unknown length
11061 11901: contig of 841 bp in length
11902 11981: gap of unknown length
11982 12648: contig of 667 bp in length
12649 12728: gap of unknown length
12729 13660: contig of 932 bp in length
13661 13740: gap of unknown length
13741 14612: contig of 872 bp in length
14613 15094: contig of 402 bp in length
15095 15174: gap of unknown length
15175 16294: contig of 1120 bp in length
16295 16374: gap of unknown length
16375 17296: contig of 922 bp in length
17297 17376: gap of unknown length
17377 18383: contig of 1007 bp in length
18384 18463: gap of unknown length
18464 19675: contig of 1212 bp in length
19676 19755: gap of unknown length
19756 21274: contig of 1519 bp in length
21275 21354: gap of unknown length
21355 22501: contig of 1147 bp in length
22502 22581: gap of unknown length
22582 23868: contig of 1287 bp in length
23869 23948: gap of unknown length
23949 24858: contig of 910 bp in length
24859 24938: gap of unknown length
24939 26046: contig of 1108 bp in length
26047 26126: gap of unknown length
26127 27036: contig of 910 bp in length
27037 27116: gap of unknown length
27117 28199: contig of 1083 bp in length
28200 28279: gap of unknown length
28280 29127: contig of 848 bp in length
29128 29207: gap of unknown length
29208 30305: contig of 1098 bp in length
30306 30385: gap of unknown length
30386 31349: contig of 964 bp in length
31350 31429: gap of unknown length
31430 32285: contig of 856 bp in length
32286 32365: gap of unknown length
32366 32749: contig of 384 bp in length
32750 32829: gap of unknown length

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galign, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneses, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 19, 2003 this sequence version replaced gi:33589945.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18260

Center clone name: 158_I_9

FEATURES

source

Location/Qualifiers

1..206000

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosomes="18"

/map="18"

/clone="RP24-158I9"

/clone_lib="RPCI-24 Male Mouse BAC"

1..1705

/note="wgs end extension"

clone end:SP6"

1706..1709

/note="clone boundary"

clone end:SP6"

site:Mb01"

complement(11831..11841)

/note="single clone coverage"

complement(22887..22896)

/note="single clone coverage"

41205..43566

/rpt_family="Lx2"

43583..44199

/rpt_family="L1M2"

complement(44209..44442)

/rpt_family="URR1B"

44735..44761

/rpt_family="(TAA)n"

44924..44951

/rpt_family="(GGGGA)n"

45069..45449

/rpt_family="Charlie7"

46341..46361

/rpt_family="AT rich"

complement(48065..48176)

/rpt_family="RMER1B"

48311..48365

/rpt_family="MLT1A0"

48514..48548

/rpt_family="MLT1A0"

48549..48654

/rpt_family="CYR11_MM"

48655..48668

/rpt_family="PBI1D9"

48669..48731

/rpt_family="MLT1A0"

repeat_region
48655..48668
/rpt_family="PBI1D9"

repeat_region
48669..48731
/rpt_family="MLT1A0"

repeat_region
48732..48842
/rpt_family="(GA)n"

repeat_region
48918..49049
/rpt_family="MLT1A0"

repeat_region
49052..49076
/rpt_family="B1_MM"

repeat_region
49141..49179
/rpt_family="(CAAA)n"

repeat_region
49843..50044
/rpt_family="AT rich"

repeat_region
50055..50200
/rpt_family="L1M4"

repeat_region
50201..50248
/rpt_family="B1_MM"

repeat_region
50249..50280
/rpt_family="A-rich"

repeat_region
50281..50419
complement(50419..50571)
/rpt_family="Tigger5"

repeat_region
50572..50637
complement(50637..50967)
/rpt_family="MTD"

repeat_region
51047..51516
/rpt_family="L1M4"

repeat_region
51752..52006
/rpt_family="B4A"

repeat_region
52067..52230
/rpt_family="ID_B1"

repeat_region
52247..52280
/rpt_family="GGA)n"

repeat_region
52281..52369
complement(52369..52419)
/rpt_family="ORR1A1"

repeat_region
52420..52454
complement(52454..54279)
/rpt_family="L1_MM"

repeat_region
54280..54275
complement(54275..55593)
/rpt_family="L1_MM"

repeat_region
55594..56221
/rpt_family="L1"

repeat_region
56222..56220
complement(56220..56444)
/rpt_family="L1"

repeat_region
56445..56683
complement(56445..56683)
/rpt_family="ORR1A2"

repeat_region
57359..57485
/rpt_family="PBI1D10"

repeat_region
57486..57837
/rpt_family="L1MA9"

repeat_region
57838..57969
/rpt_family="ID3"

repeat_region
57987..58094
/rpt_family="L1MA9"

repeat_region
58104..58301
/rpt_family="B4"

repeat_region
58302..58302
complement(58302..58696)
/rpt_family="MTA"

repeat_region
58697..58747
/rpt_family="B4"

repeat_region
58748..58858
/rpt_family="L1MA9"

repeat_region
58859..58976
/rpt_family="L1MA9"

repeat_region
58977..59055
complement(59055..59554)
/rpt_family="RLTR14"

repeat_region
61058..61101
/rpt_family="(TA)n"

repeat_region
63284..63356
/rpt_family="BC1_MM"

repeat_region
63357..63418
/rpt_family="ID_B1"

repeat_region
63419..63647
/rpt_family="URR1B"

repeat_region
63648..63840
/rpt_family="CYR11_MM"

```
Query Match      70.7%; Score 21.2; DB 10; Length 206000;
Best Local Similarity 88.5%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATCTTGCCCGCCCTTGAATTC 30
Db 124732 ATCTTGCCCGCCCTTGAATTC 124707

RESULT 13
AC141660/c
LOCUS
DEFINITION
  Apis mellifera clone CH224-54J20, WORKING DRAFT SEQUENCE, 14
ACCESSION
  AC141660.1 GI:29123844
VERSION
  HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
  Apis mellifera (honey bee)
SOURCE
  Apis mellifera
ORGANISM
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
  Apidae; Apis.
REFERENCE
  1 (bases 1 to 161580)
  Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
  Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
  Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
  Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
  Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
  Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
  Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
  Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
  Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
  Delaney,K.R., Delgado,O., Denn,A.L., Ding,X., Dinh,H.H.,
  Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
  Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
  Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
  Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
  Gorrill,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
  Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
  Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
  Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
  Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
  Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
  Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
  Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,
  Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
  Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
  Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
  Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
  Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
  Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
  Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,I.,
  Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
  Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
  Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
  Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
  Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
  Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
  Usmami,K., Vasquez,L., Vera,V., Vallalon,D., Vinson,R., Wang,Q.,
  Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
  Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
  Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
  Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 161580)
Worley K.C.
Direct Submission
Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: AMHX
Center clone name: CH224-54J20
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 158509 bases at least Q40
Consensus quality: 160165 bases at least Q30
Consensus quality: 161225 bases at least Q20
Estimated insert size: 157661; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 14 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
```

```
1 1197: contig of 1197 bp in length
* 1198 1297: gap of unknown length
* 1298 2339: contig of 1042 bp in length
* 2340 2439: gap of unknown length
* 2440 3882: contig of 1443 bp in length
* 3883 3982: gap of unknown length
* 3983 7225: contig of 3243 bp in length
* 7226 9160: contig of 1835 bp in length
* 9161 9260: gap of unknown length
* 9261 10771: contig of 1511 bp in length
* 10772 10871: gap of unknown length
* 10872 14413: contig of 3542 bp in length
* 14414 14513: gap of unknown length
* 14514 18509: contig of 3996 bp in length
* 18510 18609: gap of unknown length
* 18610 25932: contig of 7323 bp in length
* 25933 26032: gap of unknown length
* 26033 33277: contig of 7244 bp in length
* 33277 33376: gap of unknown length
* 33377 53224: contig of 19848 bp in length
* 53225 53324: gap of unknown length
* 53325 80011: contig of 26687 bp in length
* 80012 80111: gap of unknown length
* 80112 111922: contig of 31711 bp in length
* 111923 111922: gap of unknown length
* 111923 161580: contig of 49658 bp in length.
```

FEATURES

Location/Qualifiers

```
1..161580
/organism="Apis mellifera"
/mol_type="genomic DNA"
/db_xref="taxon:7460"
/cclone="CH224-54J20"
```

ORIGIN

```
Query Match      70.0%; Score 21; DB 2; Length 161580;
Best Local Similarity 82.8%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

QY 1 CTGATCTTCCCGCCCGCCCTTGAATTC 29

Db 70708 CTCGACCTCGCCCGCCCGCCCTTGAATTC 70680

RESULT 14
AC128403/c

LOCUS	AC128403	208558 bp	DNA	linear	HTG 20-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-15708, WORKING DRAFT SEQUENCE, 2				
ACCESSION	AC128403				
VERSION	AC128403.3 GI:25138164				
KEYWORDS	HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 208558) Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrat, G., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huly, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loubege, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R. A., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.				
REFERENCE	2 (bases 1 to 208558) Worley, K. C.				
AUTHORS	Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
JOURNAL	3 (bases 1 to 208558) Rat Genome Sequencing Consortium.				
REFERENCE	Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
AUTHORS	Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
JOURNAL	Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	<p>of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA</p> <p>On Nov 20, 2002 this sequence version replaced gi:23195776. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.</p> <p>Center: Genome Center Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: GYV Center clone name: CH230-15708 Assembly program: Phrap; version 0.990329 Consensus quality: 194224 bases at least Q40 Consensus quality: 196796 bases at least Q30 Consensus quality: 198381 bases at least Q20 Estimated insert size: 200025; sum-of-contigs estimation Quality coverage: 7x in Q20 bases; sum-of-contigs estimation</p> <p>* NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p> <p>* 1 192385: contig of 192385 bp in length * 192386 192485: gap of unknown length * 192486 208558: contig of 16073 bp in length.</p> <p>FEATURES source Location/Qualifiers 1..208558 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-15708" 1..1270 /note="wgs_contig" 1..600 /note="clone boundary clone_end:Sp6 site: end_sequence:BH287602" 2198..2982 /note="clone boundary clone_end:T7 site: end_sequence:BH287575"</p> <p>misc_feature misc_feature misc_feature</p> <p>Query Match 70.0%; Score 21; DB 2; Length 208558; Best Local Similarity 82.8%; Pred. No. 1.8e+02; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;</p> <p>Qy 1 CTGGATCTTGGCCCGCCCTTGATTC 29 Db 107806 CTGGATCTTGGCTCTCTTGATTC 107778</p>				
ORIGIN	<p>Query Match 70.0%; Score 21; DB 2; Length 208558; Best Local Similarity 82.8%; Pred. No. 1.8e+02; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;</p> <p>Qy 1 CTGGATCTTGGCCCGCCCTTGATTC 29 Db 107806 CTGGATCTTGGCTCTCTTGATTC 107778</p>				

JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 10, 2003 this sequence version replaced gi:24954740. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu
	----- Project Information Center project name: GHKC Center clone name: CH230-142E20 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 231949 bases at least Q40 Consensus quality: 234032 bases at least Q30 Consensus quality: 235562 bases at least Q20 Estimated insert size: 241410; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
	----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * been provided by the submitter. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. * 1 248338: contig of 248338 bp in length.
FEATURES	Location/Qualifiers 1..248338 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-142E20" 497..1140 /note="clone boundary clone_end:Sp6 site:EcoRI end_sequence:RWAXN34TVB" 10728..12176 /note="wgs contig" 242846..242895 /note="clone boundary clone_end:T7 site:EcoRI end_sequence:RWAXN34TJB"
misc_feature	source 1..248338 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-142E20" 497..1140 /note="clone boundary clone_end:Sp6 site:EcoRI end_sequence:RWAXN34TVB" 10728..12176 /note="wgs contig" 242846..242895 /note="clone boundary clone_end:T7 site:EcoRI end_sequence:RWAXN34TJB"
misc_feature	source 1..248338 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-142E20" 497..1140 /note="clone boundary clone_end:Sp6 site:EcoRI end_sequence:RWAXN34TVB" 10728..12176 /note="wgs contig" 242846..242895 /note="clone boundary clone_end:T7 site:EcoRI end_sequence:RWAXN34TJB"
misc_feature	source 1..248338 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-142E20" 497..1140 /note="clone boundary clone_end:Sp6 site:EcoRI end_sequence:RWAXN34TVB" 10728..12176 /note="wgs contig" 242846..242895 /note="clone boundary clone_end:T7 site:EcoRI end_sequence:RWAXN34TJB"
ORIGIN	Query Match 70.0%; Score 21; DB 2; Length 248338; Best Local Similarity 82.8%; Pred. No. 1.8e+02; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0; Qy 1 CTGGATCCTTCCCGCCCTTGAATTC 29 Db 30326 CTGGATCCTGGCTCTCTCTTGAATTAC 30354

RESULT 15	AC106101	248338 bp	DNA	linear	HTG 10-MAY-2003
LOCUS	AC106101				
DEFINITION	Rattus norvegicus clone CH230-142E20, WORKING DRAFT SEQUENCE.				
ACCESSION	AC106101				
VERSION	AC106101.7 GI:30521616				
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
REFERENCE	1 (bases 1 to 248338)				
AUTHORS	Muzny, D., Marie, J., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresnuhewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarpunsaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 248338)				
REFERENCE	Worley, K.C.				
AUTHORS	Direct Submission				
TITLE	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
JOURNAL	3 (bases 1 to 248338)				
REFERENCE	Rat Genome Sequencing Consortium.				
AUTHORS	Direct Submission				
TITLE	Direct Submission				

RESULT 16

AC131555 LOCUS
 DEFINITION AC131555 248825 bp DNA linear HTG 09-NOV-2002
 *** 2 unordered pieces.
 AC131555 AC131555.3 GI:24818029
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS 1 (bases 1 to 248825)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregregis, E., Geek, K., Gill, R., Grady, M., Grady, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhsu, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokelemeh, O., Okwuonu, G., Olarnpuneagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Williams, G., Willson, R., Wierzyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G., and Gibbs, R. A.

Direct Submission

Unpublished
 2 (bases 1 to 248825)

Rat Genome Sequencing Consortium.

Direct Submission

TITLE Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 248825)

AUTHORS Rat Genome Sequencing Consortium.

TITLE

JOURNAL

COMMENT

Direct Submission
 Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 9, 2002 this sequence version replaced gi:23096464.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDSA

Center clone name: CH230-10J23

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 204687 bases at least Q40

Consensus quality: 208176 bases at least Q30

Consensus quality: 210768 bases at least Q20

Estimated insert size: 215431; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 247478: contig of 247478 bp in length

* 247479 247578: gap of unknown length

* 247579 248825: contig of 1247 bp in length.

FEATURES

source

1..248825

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-10J23"

2637..3484

/note="clone boundary"

clone_end:7

site:EcoRI

end_sequence:BH306875"

34004..35851

/note="wgs_contig"

110039..185684

/note="clone boundary"

clone_end:Sp6

site:EcoRI

end_sequence:BH306876"

183163..184381

/note="wgs end extension"

clone_end:Sp6"

ORIGIN

Query Match 69.3%; Score 20.8; DB 2; Length 248825;

Best Local Similarity 91.7%; Pred. No. 2.2e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	
/gene="Evc"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
84. .1967	
/gene="Evc"	
/codon_start=1	
/product="Evc protein"	
/protein_id="AAH21464.1"	
/db_xref="GI:18204800"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
/translation="MTCTKARLQLGREALQAAPTLLVPAVLGGVLGLGLWLCGR ASHRLARLQKDRKELGSSPPAQLSDTGSQARRRORTARDADAPVCPSPISGN SLQAKGKEDSSSSSRSTYSDRILQCAFLKRVGFPPEILACESVDIDLCVCSLHKO LLQVDTALRQKHLMTIQLKACLLDFPKKPPDDELCKVKLSQKHDLLEELKGLQA RLATEMLGTGDSGVYLVADVERKELSEQLIDNMGAPKQMESIQPTLMQPKCSS SKAQFMTLGTMRIVAEGLHDSODLHVLTDLTMRGRLARLWVEFLRTQIQETK CRLAISRGLLTVQGLSGRKEELLTQOHKAFWEAERFGRFPMORGKDLVOAQ ARQAAAEALQTQEEERERELADSLTSDRGEFLKAFHEVLERQRLTSDQEGEDT WITAMAAACQELCYSTMGTFQKFDVSLFKTLPTSLPVGECSTLRQQVQEQARQ LQGMADFRFRROWGLLCLLEQDKRVWLEEGTLSTVLQRLQRDHHESTINGVLRSFGL SEE"	
Query Match 68.7%; Score 20.6; DB 10; Length 2209;	
Best Local Similarity 85.2%; Pred. No. 3.8e+02;	
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 4 GATCTTGGCCCGCCCTTGAATTCCTCC 30	
DB 1416 GATCTTGGCCCGCTGCATGATTCCTCC 1390	
RESULT 19	
BC021392/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	
/gene="Evc"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
84. .1967	
/gene="Evc"	
/codon_start=1	
/product="Evc protein"	
/protein_id="AAH21464.1"	
/db_xref="GI:18204800"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
/translation="MTCTKARLQLGREALQAAPTLLVPAVLGGVLGLGLWLCGR ASHRLARLQKDRKELGSSPPAQLSDTGSQARRRORTARDADAPVCPSPISGN SLQAKGKEDSSSSSRSTYSDRILQCAFLKRVGFPPEILACESVDIDLCVCSLHKO LLQVDTALRQKHLMTIQLKACLLDFPKKPPDDELCKVKLSQKHDLLEELKGLQA RLATEMLGTGDSGVYLVADVERKELSEQLIDNMGAPKQMESIQPTLMQPKCSS SKAQFMTLGTMRIVAEGLHDSODLHVLTDLTMRGRLARLWVEFLRTQIQETK CRLAISRGLLTVQGLSGRKEELLTQOHKAFWEAERFGRFPMORGKDLVOAQ ARQAAAEALQTQEEERERELADSLTSDRGEFLKAFHEVLERQRLTSDQEGEDT WITAMAAACQELCYSTMGTFQKFDVSLFKTLPTSLPVGECSTLRQQVQEQARQ LQGMADFRFRROWGLLCLLEQDKRVWLEEGTLSTVLQRLQRDHHESTINGVLRSFGL SEE"	
Query Match 68.7%; Score 20.6; DB 10; Length 2209;	
Best Local Similarity 85.2%; Pred. No. 3.8e+02;	
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 4 GATCTTGGCCCGCCCTTGAATTCCTCC 30	
DB 1416 GATCTTGGCCCGCTGCATGATTCCTCC 1390	
RESULT 19	
BC021392/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	
/gene="Evc"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
84. .1967	
/gene="Evc"	
/codon_start=1	
/product="Evc protein"	
/protein_id="AAH21464.1"	
/db_xref="GI:18204800"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
/translation="MTCTKARLQLGREALQAAPTLLVPAVLGGVLGLGLWLCGR ASHRLARLQKDRKELGSSPPAQLSDTGSQARRRORTARDADAPVCPSPISGN SLQAKGKEDSSSSSRSTYSDRILQCAFLKRVGFPPEILACESVDIDLCVCSLHKO LLQVDTALRQKHLMTIQLKACLLDFPKKPPDDELCKVKLSQKHDLLEELKGLQA RLATEMLGTGDSGVYLVADVERKELSEQLIDNMGAPKQMESIQPTLMQPKCSS SKAQFMTLGTMRIVAEGLHDSODLHVLTDLTMRGRLARLWVEFLRTQIQETK CRLAISRGLLTVQGLSGRKEELLTQOHKAFWEAERFGRFPMORGKDLVOAQ ARQAAAEALQTQEEERERELADSLTSDRGEFLKAFHEVLERQRLTSDQEGEDT WITAMAAACQELCYSTMGTFQKFDVSLFKTLPTSLPVGECSTLRQQVQEQARQ LQGMADFRFRROWGLLCLLEQDKRVWLEEGTLSTVLQRLQRDHHESTINGVLRSFGL SEE"	
Query Match 68.7%; Score 20.6; DB 10; Length 2209;	
Best Local Similarity 85.2%; Pred. No. 3.8e+02;	
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 4 GATCTTGGCCCGCCCTTGAATTCCTCC 30	
DB 1416 GATCTTGGCCCGCTGCATGATTCCTCC 1390	
RESULT 19	
BC021392/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	
/gene="Evc"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
84. .1967	
/gene="Evc"	
/codon_start=1	
/product="Evc protein"	
/protein_id="AAH21464.1"	
/db_xref="GI:18204800"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
/translation="MTCTKARLQLGREALQAAPTLLVPAVLGGVLGLGLWLCGR ASHRLARLQKDRKELGSSPPAQLSDTGSQARRRORTARDADAPVCPSPISGN SLQAKGKEDSSSSSRSTYSDRILQCAFLKRVGFPPEILACESVDIDLCVCSLHKO LLQVDTALRQKHLMTIQLKACLLDFPKKPPDDELCKVKLSQKHDLLEELKGLQA RLATEMLGTGDSGVYLVADVERKELSEQLIDNMGAPKQMESIQPTLMQPKCSS SKAQFMTLGTMRIVAEGLHDSODLHVLTDLTMRGRLARLWVEFLRTQIQETK CRLAISRGLLTVQGLSGRKEELLTQOHKAFWEAERFGRFPMORGKDLVOAQ ARQAAAEALQTQEEERERELADSLTSDRGEFLKAFHEVLERQRLTSDQEGEDT WITAMAAACQELCYSTMGTFQKFDVSLFKTLPTSLPVGECSTLRQQVQEQARQ LQGMADFRFRROWGLLCLLEQDKRVWLEEGTLSTVLQRLQRDHHESTINGVLRSFGL SEE"	
Query Match 68.7%; Score 20.6; DB 10; Length 2209;	
Best Local Similarity 85.2%; Pred. No. 3.8e+02;	
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 4 GATCTTGGCCCGCCCTTGAATTCCTCC 30	
DB 1416 GATCTTGGCCCGCTGCATGATTCCTCC 1390	
RESULT 19	
BC021392/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	
/gene="Evc"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
84. .1967	
/gene="Evc"	
/codon_start=1	
/product="Evc protein"	
/protein_id="AAH21464.1"	
/db_xref="GI:18204800"	
/db_xref="LocusID:59056"	
/db_xref="MGI:1890596"	
/translation="MTCTKARLQLGREALQAAPTLLVPAVLGGVLGLGLWLCGR ASHRLARLQKDRKELGSSPPAQLSDTGSQARRRORTARDADAPVCPSPISGN SLQAKGKEDSSSSSRSTYSDRILQCAFLKRVGFPPEILACESVDIDLCVCSLHKO LLQVDTALRQKHLMTIQLKACLLDFPKKPPDDELCKVKLSQKHDLLEELKGLQA RLATEMLGTGDSGVYLVADVERKELSEQLIDNMGAPKQMESIQPTLMQPKCSS SKAQFMTLGTMRIVAEGLHDSODLHVLTDLTMRGRLARLWVEFLRTQIQETK CRLAISRGLLTVQGLSGRKEELLTQOHKAFWEAERFGRFPMORGKDLVOAQ ARQAAAEALQTQEEERERELADSLTSDRGEFLKAFHEVLERQRLTSDQEGEDT WITAMAAACQELCYSTMGTFQKFDVSLFKTLPTSLPVGECSTLRQQVQEQARQ LQGMADFRFRROWGLLCLLEQDKRVWLEEGTLSTVLQRLQRDHHESTINGVLRSFGL SEE"	
Query Match 68.7%; Score 20.6; DB 10; Length 2209;	
Best Local Similarity 85.2%; Pred. No. 3.8e+02;	
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 4 GATCTTGGCCCGCCCTTGAATTCCTCC 30	
DB 1416 GATCTTGGCCCGCTGCATGATTCCTCC 1390	
RESULT 19	
BC021392/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
Young, A., Zhang, L.-H. and Green, E.D.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAK Plate: 39 Row: P Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
Location/Qualifiers	
1. .2209	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C2ECH II"	
/db_xref="taxon:10090"	
/clone="MGC:29404 IMAGE:5069004"	
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	
/clone_lib="NCI CGAP_Lu29"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .2209	

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toehiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smallus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2210)

Strausberg, R.

Direct Submission

Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapsb-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisec.nih.gov/>

Contact: nisc.mgc@hghri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granice, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

REMARK

COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAK Plate: 39 Row: p Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10947019.

FEATURES

source

Location/Qualifiers

1..2210

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CZECH II"

/db_xref="taxon:10090"

/clone="MGC:29405 IMAGE:5069005"

/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."

/clone_lib="NCI CGAP_Lu29"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

1..2210

/gene="Evc"

/db_xref="LocusID:59056"

/db_xref="MGI:1890596"

84..1967

/gene="Evc"

/codon_start=1

/product="Evc protein"

/protein_id="AAH21392.1"

gene

gene

CDS

CDS

/db_xref="GI:18204726"

/db_xref="LocusID:59056"

/db_xref="MGI:1890596"

/translation="MTCTKDARLOLQREALQAAPTLLVPAVLGGVGLGLGLWLGCR ASHARLQKODRKRLGSSPPPAQSLRDTGSOARRQRETDADDAPEVCEPSSLGN ITAFALKARVYPIQKFRPLADGSSHSLSHENLTQAAAILPHLPQPAEASPASSLG SLQAGKEDSGSSSMRSTYSDRILQCAFLRVGSFPEILACESYDIDLVCVSLHLKQ LQVDYALRQEKHLMFQILKACILIDFPKKKPDDELCKVLSKQEHLELEKGLQA RLANTMLGTGSGVSLADVERKERLESEOLIDNMGAFWKQMESIQPTLMDQPKCSS SKARQPMWLTGRMIVAEGLHDSQDLHLVTLERTMGRHLARWVEFLRTQIOBETK CLRAISLGLSLLTVQQLSGRQKEELTQQHKAPEWAEARFGRFMRGKDLVQASQ ARQAEAAALVTQTEERRSLADSQLTSDPGEFLKAFHEVLERQLRSTQDEGDET WITAMAAALCQELYSTETFKFVDSLFLKTLPEVSLPVGECETLRQQVQEQAAARQ LQADRRFRRRQWGLLCLLEQDKRWLVEEGTSLTVLQRLQRDHHESTINGVLSRFSGL SEE"

ORIGIN

Query Match 68.7%; Score 20.6; DB 10; Length 2210;

Best Local Similarity 85.2%; Pred. No. 3.8e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GATCCTTGGCCCGCCCTTGAAATTCCTCC 30

DB 1416 GATCTTTGCCCGCGTCGATGAATTCCTCC 1390

RESULT 20

ZMOPGIG/c

LOCUS Zymomonas mobilis phosphoglucose isomerase gene, complete cds. 3212 bp DNA linear BCT 26-APR-1993

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS


```
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGATCCTTGGCCCGCCCTTGAATT 27
    ||||| ||||| ||||| ||||| |||||
Db 391 CTGAGCATTGGCCCTGCGCCTTGATT 365

RESULT 21
MMU250841/c
LOCUS
DEFINITION Mus musculus mRNA for Evc protein (Ellis-van Creveld syndrome).
ACCESSION AJ250841
VERSION AJ250841.1 GI:7161865
Evc gene.
Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS Ruiz-Perez,V.L., Ide,S.E., Strom,T.M., Lorenz,B., Wilson,D.,
Woods,K., King,L., Francomano,C., Freisinger,P., Spranger,S.,
Marino,B., Dallapiccola,B., Wright,M., Meitinger,T.,
Polymopoulos,M.H. and Goodship,J.
Mutations in a new gene in Ellis-van Creveld syndrome and Weyers
acrocentral dysostosis
Nat. Genet. 24 (3), 283-286 (2000)
20164328
10700184
REFERENCE 2 (bases 1 to 4290)
Strom,T.M.
Direct Submission
Submitted (05-NOV-1999) Strom T.M., Abteilung Medizinische Genetik,
Ludwig-Maximilians-Universitaet, Goethestr. 29, Muenchen 80336,
Germany

FEATURES
source
Location/Qualifiers
1..4290
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE 569934"
/tissue type="brain cDNA, Clontech Marathon"
69..3086
/gene="Evc"
69..3086
/gene="Evc"
/note="gene mutated in Ellis van Crefeld syndrome"
/codon_start=1
/product="Evc protein"
/protein_id="CAB76567.1"
/db_xref="GI:7161866"
/db_xref="GOA:P57680"
/db_xref="UniProt/Swiss-Prot:P57680"
/translation="MTCTKDARLQGLREALQAPTLVPAVLGGVILGVLGLWLCGR
ASHLARLQKDRKRLGSSPEPASLRTGSAKARRQRTTDEDAPEVCERSLS
GNITAFKARLVYPINQFRPLADGSHSLHENTQAAILPHLPPOPAASPASS
LGLSQAQKEDSSSSSTYSDDRILQCFALRVGSPPEILACSVSDHLCVSLHL
KDLQADRTALRQEHLMFTQILKACLLDPFKKXPDDELCKQVLSQKQEHLEBKGL
QARLANTEMLGDSGYSLADVERKERELSELQIDNMGAFWKQESIOPTLMDQFKC
SSSKARQFMNTLTGMIVAEGLHDSODLHVLDTLRTWGRAHARWVEFLTQIER
TKCLRLAISRGLLELTVQQLSGRQKEILLTOHKAFWEAEERFREFWQRKDLVQA
SQARQEAALMTQSEERRSLADSLQSDGFEFLKAFHEVLRQRILRTSDQGDG
DRTITFMAALCOEYLCTMTGTFQKFDVSLFKLTPVTSPLVAECETLRQVQQA
RLQGQADRRRQGLLCLDLQQRKVMLEEGTLSTVQLQRDHHESTIHGVRFS
GLSESSRHLQGLHLLCSALRLALRGTTITALAQRSLGKRLLOELHQLAEQ
GVSPCLREHQWOLLRALEARIQEEARLEDEAQOTGLRLOQLLAEQAGLQLHM
ERVIQALLVHARNVASKRTREKDFKRLIVETVSVYVTSVNRVLQHYQAVG
KILQAEHEQLQLKTLQGERINAYKLMKKQFSDPSLESQADGTHGASQVQRML
SQQLKQDQFTTHQKRLNSRQKAEQLDQQLAEQTLQEAQETFISELSTARVPL
PNKQFNNRNGLPKPKPVTKRKKPPPREDEDLGTDDHDLALADHTTGPLSTYTSASP
PIRVHSGGRDQDSEAGDGESTSKILQGSNL"
<2963..>3044
/gene="Evc"
/note="alternative"

CDS
```

```
/codon_start=1
/product="Evc protein, short variant"
/protein_id="CAB76568.1"
/db_xref="GI:7161867"
/db_xref="GOA:P57680"
/db_xref="UniProt/Swiss-Prot:P57680"
/translation="YNLLSIASHPCPQR"

ORIGIN

Query Match 68.7%; Score 20.6; DB 10; Length 4290;
Best Local Similarity 85.2%; Pred. No. 3.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GATCCTTGGCCCGCCCTTGAATTCCC 30
    ||||| ||||| ||||| ||||| |||||
Db 1407 GATCTTTGCCCGCTGCATGAATCCC 1381

RESULT 22
AC006358/c
LOCUS
DEFINITION Homo sapiens PAC clone RP5-1127D14 from 7, complete sequence.
ACCESSION AC006358
VERSION AC006358.6 GI:15668083
Evc..
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
Wyllie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
Fewell,G.A., Deléhaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
Kalicki,J., Ozeresky,P., Bielicki,L., Scott,K., Holmes,A.,
Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
Dauphin-Kohlberg,S., Kozlowski-Reilly,A., Leonard,S., Rohlfing,T.,
Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
Strommatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
Woesner,J.P., Wendt,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
Mardis,E.R., Clifton,W., Chissoe,S.L., Mafrá,M.A., Raymond,C.,
Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadamoto,S.,
Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
The DNA sequence of human chromosome 7
Nature 424 (6945), 157-164 (2003)
22737999
12853948
REFERENCE 2 (bases 1 to 143216)
Hawrysko,C., Holmes,A. and Wohldmann,P.
The sequence of Homo sapiens PAC clone RP5-1127D14
Unpublished (2001)
JOURNAL
REFERENCE 3 (bases 1 to 143216)
Waterston,R.H.
Direct Submission
Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 143216)
Waterston,R.H.
Direct Submission
Submitted (19-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 143216)
Wilson,R.
REFERENCE
AUTHORS
```

TITLE Direct Submission
 JOURNAL Submitted (30-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Sep 19, 2001 this sequence version replaced gi:9887773.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_DJ1127D14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Bric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send malto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
 The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2.

FEATURES

source

1..143216
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP5-1127D14"
 /clone_lib="RPCI-5"
 1297..1366
 /rpt_family="L2"
 3074..3150
 /rpt_family="MaLR"
 3151..3984
 /rpt_family="ERV1"
 3885..4010
 /rpt_family="MaLR"
 4175..4287
 /rpt_family="L2"
 4386..4684
 /rpt_family="Alu"
 5430..5970
 /rpt_family="L2"
 6964..7086
 /rpt_family="MER1_type"
 8023..8334
 /rpt_family="ERV1"
 9166..9359

repeat_region
 9361..9605
 /rpt_family="MER1_type"
 repeat_region
 9725..9876
 /rpt_family="L2"
 repeat_region
 9910..9978
 /rpt_family="ERV1"
 repeat_region
 10892..11164
 /rpt_family="ERV1"
 repeat_region
 11208..11333
 /rpt_family="Alu"
 repeat_region
 11354..12349
 /rpt_family="L1"
 repeat_region
 12350..12796
 /rpt_family="L1"
 repeat_region
 12797..13068
 /rpt_family="L1"
 repeat_region
 13112..13218
 /rpt_family="Alu"
 repeat_region
 13219..13250
 /rpt_family="L1"
 repeat_region
 13251..13429
 /rpt_family="C(AAA)n"
 repeat_region
 13508..13630
 /rpt_family="Alu"
 repeat_region
 13631..13657
 /rpt_family="Alu"
 repeat_region
 13658..13825
 /rpt_family="T(AAA)n"
 repeat_region
 14031..14189
 /rpt_family="Alu"
 repeat_region
 14194..14803
 /rpt_family="L1"
 repeat_region
 14804..15109
 /rpt_family="Alu"
 repeat_region
 15110..15445
 /rpt_family="L1"
 repeat_region
 16548..16626
 /rpt_family="MER1_type"
 repeat_region
 16704..17005
 /rpt_family="Alu"
 repeat_region
 17183..17206
 /rpt_family="TTTTG)n"
 repeat_region
 17334..17637
 /rpt_family="Alu"
 repeat_region
 21282..21369
 /rpt_family="L2"
 repeat_region
 22270..22567
 /rpt_family="Alu"
 repeat_region
 22753..23000
 /rpt_family="MIR"
 repeat_region
 23009..23128
 /rpt_family="L2"
 repeat_region
 23230..23528
 /rpt_family="Alu"
 repeat_region
 23667..23871
 /rpt_family="Alu"
 repeat_region
 24365..24764
 /rpt_family="ERV1"
 repeat_region
 25154..25485
 /rpt_family="Alu"
 repeat_region
 26039..26533
 /rpt_family="ERV1"
 repeat_region
 27567..27700
 /rpt_family="Alu"
 repeat_region
 27701..27816
 /rpt_family="MIR"
 repeat_region
 27947..28112
 /rpt_family="MIR"
 repeat_region
 29769..30078
 /rpt_family="Alu"
 repeat_region
 30094..30366
 /rpt_family="Alu"

```

repeat_region 30877..30917
Query Match      68.7%; Score 20.6; DB 9; Length 143216;
Best Local Similarity 85.2%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCTTGAAAT 27
Db 135826 CTGAGCCGAGCCCGCCCTTGGAAT 135800

RESULT 23
AC084867
LOCUS
DEFINITION
  Homo sapiens chromosome 7 clone RP11-113A18, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
ACCESSION
  AC084867
VERSION
  AC084867.1 GI:11323402
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 165576)
REFERENCE
  Waterston,R.H.
  The sequence of Homo sapiens clone
  Unpublished
  2 (bases 1 to 165576)
  Waterston,R.H.
  Direct Submission
  Submitted (24-NOV-2000) Genome Sequencing Center, Washington
  University School of Medicine, 444 Forest Park Parkway, St. Louis,
  MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0113A18
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primed ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151987 bases at least Q40
Consensus quality: 156444 bases at least Q30
Consensus quality: 158675 bases at least Q20
Insert size: 171000; agarose-fp
Quality coverage: 4.41 in Q20 bases; agarose-fp
Quality coverage: 4.67 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2758: contig of 2758 bp in length
* 2759 2858: gap of unknown length
* 2859 7421: contig of 4563 bp in length
* 7422 7522: gap of unknown length
* 7522 10502: contig of 2981 bp in length
* 10503 10602: gap of unknown length
* 10603 15682: contig of 5080 bp in length
* 15683 15782: gap of unknown length
* 15783 22462: contig of 6680 bp in length
* 22463 22562: gap of unknown length
* 22563 28905: contig of 6343 bp in length

FEATURES
            source
            Location/Qualifiers
                1..165576
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="7"
                /clone="RP11-113A18"
                1..2758
                /note="assembly_name:Contig10"
                2859..7421
                /note="assembly_name:Contig11"
                7522..10502
                /note="assembly_name:Contig12"
                10603..15682
                /note="assembly_name:Contig13"
                15783..22462
                /note="assembly_name:Contig14"
                clone_end:SP6
                vector_side:left
                22563..28905
                /note="assembly_name:Contig15"
                29006..42474
                /note="assembly_name:Contig16"
                42575..52168
                /note="assembly_name:Contig17"
                52269..64465
                /note="assembly_name:Contig18"
                64566..83293
                /note="assembly_name:Contig19"
                83394..105224
                /note="assembly_name:Contig20"
                105325..153100
                /note="assembly_name:Contig21"
                153201..154699
                /note="assembly_name:Contig3"
                154800..156690
                /note="assembly_name:Contig5"
                156791..159329
                /note="assembly_name:Contig6"
                159430..161813
                /note="assembly_name:Contig7"
                161914..164135
                /note="assembly_name:Contig8"
                164236..165576
                /note="assembly_name:Contig9"

ORIGIN
Query Match      68.7%; Score 20.6; DB 2; Length 165576;
Best Local Similarity 85.2%; Pred. No. 2.8e+02;

```


VECTOR: pBACe3.6.
 Location/Qualifiers
 1. 73066
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosomes="2"
 /clone_lib="RP23-171K6"
 /clone_lib="RPCI-23"

ORIGIN
 Query Match 68.0%; Score 20.4; DB 10; Length 73066;
 Best Local Similarity 80.0%; Pred. No. 3.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGATCTTGGCCCGCCCTTGAATCCC 30
 |||||
 Db 39000 CTGATCTTCCAGCCCGCTGACTCCC 38971

RESULT 26
 AC099796 132033 bp DNA linear PRI 28-DEC-2002
 LOCUS Homo sapiens chromosome 1 clone RP5-866L20, complete sequence.
 DEFINITION AC099796 AL162404
 ACCESSION AC099796.2 GI:27413195
 VERSION HTG.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 132033)
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
 Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
 Haugen,B.D.
 Direct Submission
 Unpublished
 2 (bases 1 to 132033)
 Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,B.D.
 Direct Submission
 Submitted (21-NOV-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 132033)
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
 Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
 Haugen,B.D.
 Direct Submission
 Submitted (28-DEC-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Dec 28, 2002 this sequence version replaced gi:17027312.
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: wgchgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP5-866L20 (sc0841)
 ----- Summary Statistics
 Sequencing vector: plasmid; 58% of reads
 Sequencing vector: plasmid; 42% of reads
 Chemistry: Dye-terminator ET; 38% of reads
 Chemistry: Dye-terminator Big Dye; 62% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 131632 bases at least Q40
 Consensus quality: 131970 bases at least Q30
 Consensus quality: 132027 bases at least Q20
 Insert size: 132033; sum-of-contigs
 Quality coverage: 8.4x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': RP4-705F19 AL035415, 2428-bp overlap

3': RP11-240D10 AL590093, 2000-bp overlap.
 This is a partial submission, omitting 16999 bp of the 5' overlap.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII				EcoRI				BglII			
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt				
1123	1101	11398	11214	7830	7678						
449	<800	2184	2165	5671	5644						
512	<800	8065	8029	11476	11227						
2814	2813	486	<800	31	<800						
1247	1199	106	<800	9846	10010						
4695	4665	2759	2797	2554	2491						
3373	3396	691	<800	3354	3351						
3371	3396	1119	1129	5799	5849						
1726	1720	4445	4410	6667	6703						
7553	7674	16360	16101	7445	7678						
7059	7096	1176	1129	1444	1417						
12237	11789	761	<800	955	981						
1348	1329	3290	3383	227	<800						
145	<800	18041	18453	1841	1843						
15	<800	1543	1447	2713	2764						
703	<800	6296	6386	9714	9566						
4893	4872	3669	3631	4665	4634						
978	987	734	704	7425	7396						

TITLE Direct Submission
JOURNAL Submitted (01-SEP-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 1, 2004 this sequence version replaced gi:50199041.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L25532
Center clone name: 340_J_12

FEATURES
source
1..149118
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP24-340J12"
/clone_lib="RPC1-24 Male Mouse BAC"
1..4
/note="clone boundary
clone end:SP6
site:Mbol"
misc_feature
141..364
/rpt_family="B3A"
357..491
/rpt_family="B1_MM"
4349..4375
/rpt_family="(GGGA)n"
4402..4440
/rpt_family="(CAGA)n"
5770..5795
/rpt_family="(CA)n"
10212..10242
/rpt_family="(TG)n"
13257..13304
/rpt_family="CT-rich"
14786..15189
/rpt_family="MTD"
15555..15595
/rpt_family="(CA)n"
15725..15826
/rpt_family="PB1D9"
15943..16072
/rpt_family="L1M4"
16075..16196
/rpt_family="RSINE1"
16195..16221
/rpt_family="(CA)n"
complement(20382..20526)
/rpt_family="B1_MM"
20635..20666
/rpt_family="(A)n"
21017..21134
/rpt_family="ID_B1"
21164..21272
/rpt_family="ID_B1"
22345..22406
/rpt_family="AT-rich"
complement(22407..22557)
/rpt_family="B1_MM"
complement(22559..22766)
/rpt_family="B3A"
complement(22782..22906)
/rpt_family="B3A"
22907..23056
/rpt_family="B1_MM"

repeat_region complement(23057..23116)
/rpt_family="B3A"
29721..29796
/rpt_family="(A)n"
complement(31059..31193)
/rpt_family="B1_MM"
complement(31500..31639)
/rpt_family="B1_MM"
complement(31509..31713)
/rpt_family="ID_B1"
31801..31996
/rpt_family="B2_MM1"
32041..33336
/rpt_family="ORR1A-int"
33337..33669
/rpt_family="ORR1A3"
34036..34140
/rpt_family="RSINE1"
34141..34291
/rpt_family="B1_MM"
34292..34345
/rpt_family="RSINE1"
34513..34685
/rpt_family="B4"
34557..34689
/rpt_family="B1_MM"
34693..34839
/rpt_family="B1_MM"
complement(35350..35555)
/rpt_family="B3"
35716..35793
/rpt_family="MLT1A0"
35794..35982
/rpt_family="B2_MM1"
35983..36028
/rpt_family="MLT1A0"
complement(36094..36219)
/rpt_family="B1_MM"
36261..36570
/rpt_family="MT2A"
complement(36816..36987)
Query Match 68.0%; Score 20.4; DB 10; Length 149118;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCTTCCCGCCCTTGAAATCCC 30
DB 104968 CTGGAGCCCTGGCCCTCTTGAAATTC 104939
RESULT 28
AC004383
LOCUS AC004383 156461 bp DNA linear PRI 08-MAR-1998
DEFINITION Human Chromosome X clone bxXD187, complete sequence.
ACCESSION AC004383
VERSION AC004383.1 GI:2944111
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156461)
AUTHORS Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
Mazzarella, R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 156461)
AUTHORS Brownstein, B.H., States, D.J. and Mazzarella, R.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA

COMMENT

Current status of this project is available at:
'http://www.ibc.wustl.edu/cgm/seq_projects.html',
Submitted by:
Elison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: elison@genseq.aplbio.com
and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu
and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

FEATURES

source
1..156461
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="BXVD187"

ORIGIN

Query Match 68.0%; Score 20.4; DB 9; Length 156461;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTCGCCCGCCCGCCCTTGATTC 30
||||| ||||| ||||| ||||| ||||| |||||
DB 141617 CTGGATCTTCGCCCGCCCGCCCGATTC 141646

RESULT 29

AC102506/c
LOCUS 163311 bp DNA linear HTG 18-JUN-2004
DEFINITION Mus musculus chromosome 1 clone RP24-139E15 map 1, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.

ACCESSION AC102506 GI:45642843
VERSION AC102506.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPIN.
SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 163311)

TITLE Mus musculus chromosome 1, clone RP24-139E15
JOURNAL Unpublished

2 (bases 1 to 163311)

REFERENCE
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 163311)

REFERENCE
Birken, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataranan, V.S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.

Direct Submission

Submitted (18-JUN-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 23, 2004 this sequence version replaced gi:29135684.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L18931

Center clone name: 139_E_15

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 20812: contig of 20812 bp in length
* 20813 20912: gap of 100 bp
* 20913 64207: contig of 43295 bp in length
* 64208 64307: gap of 100 bp
* 64308 163311: contig of 99004 bp in length.

Location/Qualifiers

1..163311
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP24-139E15"

FEATURES

source

/clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match 68.0%; Score 20.4; DB 2; Length 163311;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCCTTGGCCCGCCCTTGAATCC 30
 |||||
 Db 79721 CTGGATCCCTTGGCCCGCCCTTGGCCATCC 79692
 |||||

RESULT 30

AC025990/c 166549 bp DNA linear HTG 03-MAY-2000
 LOCUS Homo sapiens chromosome 1 clone RP11-277A12 map 1, WORKING DRAFT
 DEFINITION SEQUENCE, 30 unordered pieces.

AC025990

AC025990.2 GI:7684488

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 166549)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Baetien,V., Beda,F.,

Boguslavsky,L., Bouckgatter,B., Brown,A., Burkett,G.,

Campolano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galgan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Strange-Thomann,N., Srojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141 USA

On May 3, 2000 this sequence version replaced gi:7263208.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8019

Center clone name: 277_A12

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 148491 bases at least Q40

Consensus quality: 156797 bases at least Q30

Consensus quality: 160643 bases at least Q20

Insert size: 171000; agarose-fp

FEATURES

source

Insert size: 163649; sum-of-contigs
 Quality coverage: 4.1 in Q20 bases; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 30 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1034: contig of 1034 bp in length
 * 1035 1134: gap of 100 bp
 * 1135 2525: contig of 1391 bp in length
 * 2526 2625: gap of 100 bp
 * 2626 4050: contig of 1425 bp in length
 * 4051 4150: gap of 100 bp
 * 4151 5525: contig of 1375 bp in length
 * 5526 7121: contig of 1496 bp in length
 * 7122 8455: contig of 1134 bp in length
 * 8456 10273: contig of 1818 bp in length
 * 10274 10373: gap of 100 bp
 * 10374 12187: contig of 1814 bp in length
 * 12188 12287: gap of 100 bp
 * 12288 13713: contig of 1426 bp in length
 * 13714 13813: gap of 100 bp
 * 13814 16806: contig of 2993 bp in length
 * 16807 16906: gap of 100 bp
 * 16907 17920: contig of 1014 bp in length
 * 17921 18020: gap of 100 bp
 * 18021 21078: contig of 3058 bp in length
 * 21079 21178: gap of 100 bp
 * 21179 24583: contig of 3405 bp in length
 * 24584 24683: gap of 100 bp
 * 24684 29247: contig of 4564 bp in length
 * 29248 29347: gap of 100 bp
 * 29348 34482: contig of 5135 bp in length
 * 34483 34582: gap of 100 bp
 * 34583 37621: contig of 3039 bp in length
 * 37622 37721: gap of 100 bp
 * 37722 43206: contig of 5485 bp in length
 * 43207 43306: gap of 100 bp
 * 43307 45997: contig of 2691 bp in length
 * 45998 46097: gap of 100 bp
 * 46098 52449: contig of 6352 bp in length
 * 52450 52549: gap of 100 bp
 * 52550 57930: contig of 5381 bp in length
 * 57931 58030: gap of 100 bp
 * 58031 64917: contig of 6887 bp in length
 * 64918 65017: gap of 100 bp
 * 65018 74493: contig of 9476 bp in length
 * 74494 74593: gap of 100 bp
 * 74594 81921: contig of 7328 bp in length
 * 81922 82021: gap of 100 bp
 * 82022 89795: contig of 7774 bp in length
 * 89796 89895: gap of 100 bp
 * 89896 100646: contig of 10751 bp in length
 * 100647 100746: gap of 100 bp
 * 100747 109185: contig of 8439 bp in length
 * 109186 109285: gap of 100 bp
 * 109286 120571: contig of 11286 bp in length
 * 120572 120671: gap of 100 bp
 * 120672 130560: contig of 9889 bp in length
 * 130561 130660: gap of 100 bp
 * 130661 145409: contig of 14749 bp in length
 * 145410 145509: gap of 100 bp
 * 145510 166549: contig of 21040 bp in length.
 Location/Qualifiers
 1..166549


```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17672
Center clone name: 349 H 12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173711 bases at least Q40
Consensus quality: 175255 bases at least Q30
Consensus quality: 175819 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 176018; sum-of-contigs
Quality coverage: 12.1 in Q20 bases; agarose-fp
Quality coverage: 11.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 11171: contig of 11171 bp in length
* 11172: gap of 100 bp
* 12172: contig of 1079 bp in length
* 12351: gap of 100 bp
* 12431: contig of 1115 bp in length
* 13585: gap of 100 bp
* 13665: contig of 1422 bp in length
* 15087: gap of 100 bp
* 15188: contig of 1304 bp in length
* 16491: gap of 100 bp
* 16492: gap of 100 bp
* 16592: contig of 118509 bp in length
* 135101: gap of 100 bp
* 135201: contig of 1668 bp in length
* 136868: gap of 100 bp
* 136869: contig of 1353 bp in length
* 138321: gap of 100 bp
* 138322: contig of 2475 bp in length
* 138422: contig of 100 bp
* 140897: contig of 35064 bp in length
* 140937: gap of 100 bp
* 176061: contig of 858 bp in length.
* 176161 177018: contig of 858 bp in length.
URES      Location/Qualifiers
source     1..177018
           /organism="Mus musculus"
           /mol_type="Genomic DNA"
           /db_xref="taxon:10090"
           /clone="RP23-349H12"
           /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1..11171
            /note="assembly_fragment"
            clone_end:SP6
vector_side:left
11272..12350
/note="assembly_fragment"
12451..13585
/note="assembly_fragment"
13666..15087
/note="assembly_fragment"
15188..16491
/note="assembly_fragment"
16592..135100
/note="assembly_fragment"
135201..136868

```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source

```

1..178620
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="8"
/map="8"
/clone="RP24-11201"
/clone_lib="RPCI-24"
221..424
/rpt_family="L1"
886..962
/rpt_family="ID"
1994..2120
/rpt_family="L1"
2628..2871
/rpt_family="B4"
6005..6098
/rpt_family="MIR"
7444..7653
/rpt_family="MER1_type"
7980..8175
/rpt_family="B4"
8109..8195
/rpt_family="B4"
9379..9770
/rpt_family="MaLR"
10306..10473
/rpt_family="MER2_type"
10510..10649
/rpt_family="Alu"
10724..11075
/rpt_family="MER2_type"
11237..11290
/rpt_family="ERV1"
11420..11932
/rpt_family="L1"
12498..12563
/rpt_family="L2"
12882..13087
/rpt_family="B2"
14149..14252
/rpt_family="Alu"
14721..14924
/rpt_family="B2"
15429..15516
/rpt_family="RMER3"
15584..16064
/rpt_family="ERVK"
16070..16625
/rpt_family="ERVK"
16658..16710
/rpt_family="ERVK"

```

```

repeat_region 16720..16971
/rpt_family="ERVK"
repeat_region 16967..17091
/rpt_family="ERVK"
repeat_region 17094..17731
/rpt_family="ERVK"
repeat_region 17730..18517
/rpt_family="ERVK"
repeat_region 18518..18998
/rpt_family="ERVK"
repeat_region 19521..19869
/rpt_family="ERVK"
repeat_region 20747..20785
/rpt_family="ID"
repeat_region 20889..20964
/rpt_family="ID"
repeat_region 22207..22318
/rpt_family="Alu"
repeat_region 22319..22347
/rpt_family="B4"
repeat_region 22774..22924
/rpt_family="B4"
repeat_region 23116..23349
/rpt_family="L1"
repeat_region 23769..23884
/rpt_family="MIR"
repeat_region 24123..24331
/rpt_family="L1"
repeat_region 24356..24758
/rpt_family="MaLR"
repeat_region 24763..24900
/rpt_family="Alu"
repeat_region 24980..25125
/rpt_family="L1"
repeat_region 25494..25546
/rpt_family="L1"
repeat_region 25693..25843
/rpt_family="L1"
repeat_region 26367..26727
/rpt_family="L1"
repeat_region 33524..33614
/rpt_family="L2"
repeat_region 34301..34352
/rpt_family="B4"
repeat_region 34419..34490
/rpt_family="B4"
repeat_region 34653..34888
/rpt_family="ERVK"
repeat_region 36144..36350
/rpt_family="ERVK"
repeat_region 38320..38378
/rpt_family="MIR"
repeat_region 38524..38598
/rpt_family="ID"
repeat_region 38529..38597
/product="tRNA-Ala"
/rnote="Likely pseudogene (HMM Sc=28.18 / Sec struct
Sc=-6.76)"
repeat_region 38685..38786
/rpt_family="B4"
repeat_region 41214..41432
/rpt_family="MIR"
repeat_region 41907..42131
/rpt_family="B4"
repeat_region 42096..42177
/rpt_family="MaLR"
repeat_region 44358..44496
/rpt_family="B4"
repeat_region 45096..45278
/rpt_family="B2"
repeat_region 45099..45174
/product="tRNA-Ser"
/rnote="Likely pseudogene (HMM Sc=25.51 / Sec struct
trna

```

repeat_region
 repeat_region
 SC=-4.30"
 46778..46911
 /rpt_family="L1"
 47278..47332
 /rpt_family="ERV1"

Query Match 68.0%; Score 20.4; DB 10; Length 178620;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCTTGGCCCGCCCTTGAATTCCTCC 30
 DB 13707 CTGGCCATTGCAGCCCGCTGAATTCCTCC 13736

RESULT 33
 AC010965/c
 LOCUS
 DEFINITION Mus musculus clone RP23-479L21, *** SEQUENCING IN PROGRESS ***, 2
 ordered pieces.
 AC010965
 HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 186208)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-479L21

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 186208)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission

Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 186208)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
 Peterson, K.H., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission

Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 19, 2002 this sequence version replaced gi:11496385.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L908
 Center clone name: 479_L_21

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 176982: contig of 176982 bp in length
 * 176983 177082: gap of 100 bp
 * 177083 186208: contig of 9126 bp in length.

FEATURES

Location/Qualifiers
 1..186208
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-479L21"
 /clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 68.0%; Score 20.4; DB 2; Length 186208;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 CTGGATCTTGGCCCGCCCTTGAATTCCTCC 30
 DB 148151 CTGAATGCTTTCCAGCCCGCTGACTTCCC 148122

RESULT 34

PTB083D21/c
 LOCUS
 DEFINITION Pan troglodytes chromosome X BAC PTB-083D21, complete sequence.
 ACCESSION CR380839
 VERSION CR380839.1 GI:45752336

KEYWORDS HTG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (bases 1 to 186291)
 Martinez-Arias, R., Conrad, A., Loehner, T.H., Ludewig, M.,
 Scharfe, M., Schindewolf, C., Schrader, F., Thies, S. and Bloeker, H.
 Direct Submission
 Submitted (25-MAR-2004) GBF, Dept. of Genome Analysis, Mascheroder
 Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de
 All annotations in this database entry are developed by
 computational tools. It is therefore not explicitly noted in the
 feature lines that evidence is not experimental.

COMMENT

----- Genome Center
 Center: GBF, Braunschweig
 Center code: GBF
 Web site: http://genome.gbf.de/
 Contact: info.genome@gbf.de
 ----- Project Information
 Center project name:
 Center clone name: PTB-083D21
 ----- Summary Statistics
 Sequencing vector: ###;
 Chemistry: Dye-terminator-amersham: ## of reads

[illegible]

repeat_region	/rpt_family="L1"	8932..8991
repeat_region	/rpt_family="ID"	9647..9847
repeat_region	/rpt_family="MER2_type"	9954..10191
repeat_region	/rpt_family="L1"	10212..10309
repeat_region	/rpt_family="Alu"	10685..11184
repeat_region	/rpt_family="MaLR"	11563..11845
repeat_region	/rpt_family="L1"	11846..11918
repeat_region	/rpt_family="MaLR"	12486..12548
repeat_region	/rpt_family="MaLR"	12550..12604
repeat_region	/rpt_family="L1"	12605..12922
repeat_region	/rpt_family="MaLR"	12926..13035
repeat_region	/rpt_family="L1"	13731..13950
repeat_region	/rpt_family="L1"	15430..15917
repeat_region	/rpt_family="L1"	16861..16950
repeat_region	/rpt_family="MER1_type"	17060..17222
repeat_region	/rpt_family="MER1_type"	17925..17977
repeat_region	/rpt_family="ERV1"	18229..18398
repeat_region	/rpt_family="B4"	21073..21223
repeat_region	/rpt_family="L1"	23859..24006
repeat_region	/rpt_family="Alu"	25443..25606
repeat_region	/rpt_family="MER1_type"	26780..26882
repeat_region	/rpt_family="Alu"	27931..28077
repeat_region	/rpt_family="MaLR"	29275..29436
repeat_region	/rpt_family="MER1_type"	30361..30460
repeat_region	/rpt_family="Alu"	31549..31674
repeat_region	/rpt_family="MIR"	32061..32126
repeat_region	/rpt_family="ID"	32772..33496
repeat_region	/rpt_family="L1"	33598..33980
repeat_region	/rpt_family="L1"	33977..34079
repeat_region	/rpt_family="L1"	34253..34404
repeat_region	/rpt_family="MaLR"	34408..34476
repeat_region	/rpt_family="L1"	34465..35614
repeat_region	/rpt_family="L1"	35617..35878
repeat_region	/rpt_family="MaLR"	35923..36021
repeat_region	/rpt_family="Alu"	36022..36242
repeat_region	/rpt_family="L1"	36333..36411
repeat_region	/rpt_family="MaLR"	

```

repeat_region 36396..36546 /rpt_family="MaLR"
repeat_region 36552..36747 /rpt_family="B2"
repeat_region 37199..37282 /rpt_family="L1"
repeat_region 38061..38417 /rpt_family="MaLR"
repeat_region 41309..41555 /rpt_family="L1"
repeat_region 42632..42874 /rpt_family="B4"
repeat_region 43295..43376 /rpt_family="B2"
repeat_region 43457..43629 /rpt_family="B2"
repeat_region 45735..45749 /note="sequence derived from one plasmid subclone."
unsure 50417..50600 /rpt_family="B4"
repeat_region

```

```

Query Match      68.0%; Score 20.4; DB 10; Length 191190;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

y      1 CTGGATCCTTGGCCCGCCCTTGAAATCCC 30
      ||||| ||||| ||||| |||||
b      150690 CTGGTGTCTTGTCGCGCCCTTGACCCCC 150719

RESULT 36
K105989/c
OCUS
DEFINITION
Mus musculus chromosome 8, clone RP24-95N3, complete sequence.
linear ROD 29-APR-2004

```

TITLE Direct Submission
JOURNAL Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 194387)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Zaimoun, O., Zenger, E., Zimmer, S. and Zody, M.


```

/rpt_family="RMER17A2"
repeat_region 17008..17084
/rpt_family="L2"
repeat_region 17678..17699
/rpt_family="(TA)n"
repeat_region complement(17700..17847)
/rpt_family="B1P"
repeat_region 18206..18238
/rpt_family="(TTTTTG)n"
repeat_region complement(18285..18408)
/rpt_family="B1 MM"
repeat_region complement(20260..20429)
/rpt_family="Lx6"

Query Match 68.0%; Score 20.4; DB 10; Length 194387;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCGCCCTTGAATTCCC 30
Db 66733 CTGGCCATTGCCAGGCCCGCGTGAATTCCC 66704

RESULT 37
AC091762
LOCUS AC091762 204734 bp DNA linear ROD 20-JUN-2003
DEFINITION Mus musculus clone rp23-151n4 map 2 strain C57BL/6J, complete
sequence.
ACCESSION AC091762
VERSION AC091762.27 GI:32129349
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 204734)
AUTHORS Jiang, X., Song, L. and Roe, B.A.
TITLE Mus musculus Chromosome 10 BAC Clone rp23-151n4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204734)
AUTHORS Jiang, X., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 204734)
AUTHORS Jiang, X., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 204734)
AUTHORS Jiang, X., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jun 20, 2003 this sequence version replaced gi:31581643.

----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR
-----
FEATURES
Source
1..204734
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="2"
/clone="rp23-151n4"
/clone_lib="RPCL - 23 Female (C57BL/6J) Mouse BAC Library"

ORIGIN
***, 2 ordered pieces.

Query Match 68.0%; Score 20.4; DB 10; Length 194387;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCGCCCTTGAATTCCC 30
Db 66733 CTGGCCATTGCCAGGCCCGCGTGAATTCCC 66704

RESULT 38
AC147615
LOCUS AC147615 205962 bp DNA linear ROD 07-AUG-2004
DEFINITION Mus musculus chromosome 10 clone RP23-408N24, complete sequence.
ACCESSION AC147615
VERSION AC147615.3 GI:51036445
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205962)
AUTHORS Wilson, R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205962)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 205962)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 205962)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 7, 2004 this sequence version replaced gi:49406112.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----
Project Information
-----
Center project name: M_BA0408N24
-----
Location/Qualifiers
Source
1..205962
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10"
/clone="RP23-408N24"

ORIGIN
Query Match 68.0%; Score 20.4; DB 10; Length 205962;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGCCCGCCCGCCCTTGAATTCCC 30
Db 126929 CTGGTTGTTGCTCCGCCCTTGACCCCC 126958

RESULT 39
AC004387
LOCUS AC004387 206156 bp DNA linear HTG 13-JUN-2002
DEFINITION Homo sapiens chromosome X clone bWXD173, *** SEQUENCING IN PROGRESS
***, 2 ordered pieces.

```

```

ACCESSION AC004387.2 GI:21405643
VERSION HTG; HTGS_PHASE2.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE
AUTHORS Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
          Mazzarella, R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 206156)
AUTHORS Brownstein, B.H., States, D.J. and Mazzarella, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1998) Center for Genetics in Medicine, Box 8232,
          Washington University School of Medicine, 4566 Scott Avenue, St.
          Louis, MO 63110, USA
COMMENT On Jun 13, 2002 this sequence version replaced gi:2944106.
          Current status of this project is available at:
          'http://www.ibt.wustl.edu/cgm/seq_projects.html'
          Submitted by:
          Eilison Chen,
          Advanced Center for Genetic Technology,
          Applied Biosystems Division of Perlin Elmer Corp.,
          850 Lincoln Center Drive,
          Foster City, CA 94404 USA
          e-mail: elison@genseq.apldbio.com
          and
          Buddy Brownstein,
          Center for Genetics in Medicine,
          Washington University School of Medicine, Box 8232
          4566 Scott Avenue,
          St. Louis, MO 63110, USA
          e-mail: buddy@genetics.wustl.edu
          and
          David J. States,
          Institute for Biomedical Computing
          Washington University in St. Louis
          700 South Euclid Ave.
          St. Louis, MO 63108 USA
          e-mail: states@ibc.wustl.edu.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          * 1 195298: contig of 195298 bp in length
          * 195299 195398: gap of 100 bp
          * 195399 206156: contig of 10758 bp in length.
          FEATURES
          source
            1..206156
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="X"
              /clone="bWDXD173"
          ORIGIN
            Query Match 68.0%; Score 20.4; DB 2; Length 206156;
            Best Local Similarity 80.0%; Pred. No. 3.4e+02;
            Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
            QY 1 CTGGATTCTTCCCGCCCGCCCTTGAATTCCTCC 30
              ||||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 86494 CTGGATTCTTCCCGCCCGCCCGGTTATCCC 86523
RESULT 40
AL590875 211025 bp DNA linear HTG 02-OCT-2001
LOCUS Homo sapiens chromosome X clone RP11-671P4 map q26.1-26.3, 3
DEFINITION unordered pieces.
ACCESSION AL590875
VERSION AL590875.9 GI:15912458
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Bray-Allen, S.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
          requests: clonerequest@sanger.ac.uk
          On Oct 3, 2001 this sequence version replaced gi:14349079.
COMMENT ----- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ----- Project Information
          Center project name: BM671P4
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; L08752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 190476 bases at least Q40
          Consensus quality: 191400 bases at least Q30
          Consensus quality: 192001 bases at least Q20
          Insert size: 210825; sum-of-contigs
          Insert size: 195205; 0.2% error; agarose-fp
          Quality coverage: 6.29x in Q20 bases; sum-of-contigs Quality
          coverage: 7.20x in Q20 bases; agarose-fp
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence.
          * as soon as it is available and the accession number will
          * be preserved.
          * 1 10462: contig of 10462 bp in length
          * 10463 10562: gap of 100 bp
          * 10563 16055: contig of 5493 bp in length
          * 16056 16155: gap of 100 bp
          * 16156 211025: contig of 194870 bp in length.
          FEATURES
          source
            1..211025
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="X"
              /map="q26.1-26.3"
              /clone="RP11-671P4"
              /clone.lib="RPC1-11.3"
              1..10462
                /note="assembly_fragment:00031"
                10563..16055
                  /note="assembly_fragment:01394"
                  16156..211025
                    /note="assembly_fragment:04589"
                    clone_end:SP6
                    vector_side:right
            misc_feature
              /note="assembly_fragment:00031"
              10563..16055
            misc_feature
              /note="assembly_fragment:01394"
              16156..211025
            misc_feature
              /note="assembly_fragment:04589"
              clone_end:SP6
              vector_side:right
          ORIGIN

```

Query Match	68.0%; Score 20.4; DB 2; Length 211025;	
Best Local Similarity	80.0%; Pred. No. 3.4e+02;	
Matches	24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	1 CTGATCTTCGCCCGCCCGCCCTTGAAATTCCTCC 30	
Db	196183 CTGATCTTCGCCCGCCCGCCCGATTCCTCC 196212	
RESULT 41		
AC134155		
LOCUS	Rattus norvegicus clone CH230-180114, *** SEQUENCING IN PROGRESS	
DEFINITION	***, 5 unordered pieces.	
ACCESSION	AC134155	
VERSION	AC134155.2 GI:25137725	
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.	
SOURCE	Rattus norvegicus (Norway rat)	
ORGANISM	Rattus norvegicus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE	1 (bases 1 to 231160)	
AUTHORS	Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gegeorge, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowig, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzen, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.	
Direct Submission		
Unpublished		
2 (bases 1 to 231160)		
Rat Genome Sequencing Consortium.		
Direct Submission		
Submitted (23-SEP-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
3 (bases 1 to 231160)		
Rat Genome Sequencing Consortium.		
Direct Submission		
Submitted (20-NOV-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On Nov 20, 2002	this sequence version replaced gi:23269748.	
The sequence in this assembly	is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
Center: Baylor College of Medicine		
Center code: BCM		
Web site: http://www.hgsc.bcm.tmc.edu/		
Contact: hgsc-help@bcm.tmc.edu		
----- Project Information		
Center project name: GVHL		
Center clone name: CH230-180114		
----- Summary Statistics		
Assembly program: Phrap; version 0.990329		
Consensus quality: 197731 bases at least Q40		
Consensus quality: 201237 bases at least Q30		
Consensus quality: 203730 bases at least Q20		
Estimated insert size: 198394; sum-of-contigs estimation		
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation		

* NOTE: Estimated insert size may differ from sequence length		
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).		
* NOTE: This is a 'working draft' sequence. It currently		
* consists of 5 contigs. The true order of the pieces		
* is not known and their order in this sequence record is		
* arbitrary. Gaps between the contigs are represented as		
* runs of N, but the exact sizes of the gaps are unknown.		
* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
* 1 11920: contig of 11920 bp in length		
* 11921 12020: gap of unknown length		
* 12021 62724: contig of 50704 bp in length		
* 62725 62824: gap of unknown length		
* 62825 135634: contig of 72810 bp in length		
* 135635 135734: gap of unknown length		
* 135735 231905: contig of 96171 bp in length		
* 231906 232005: gap of unknown length		
* 232006 233160: contig of 1155 bp in length.		
Location/Qualifiers		
1. 233160		
/organism="Rattus norvegicus"		
/mol_type="genomic DNA"		
/db_xref="taxon:10116"		
/clone="CH230-180114"		
12021. 15322		
/note="wgs_contig"		
192309. 193315		
/note="wgs_contig"		
misc_feature		
misc_feature		
ORIGIN		

TITLE

TITLE
JOURNAL
COMMENT

Direct Submission
 Submitted (23-SEP-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 23, 2004 this sequence version replaced gi:22381891.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@broad.mit.edu
 ----- Project Information
 Center project name: L20060
 Center clone name: 303_N_12

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	17526:	contig of 17526 bp in length
*	17627	17626: gap of unknown length
*	17627	22764: contig of 5138 bp in length
*	22765	22864: gap of unknown length
*	22865	31130: contig of 8266 bp in length
*	31131	31230: gap of unknown length
*	31231	46325: contig of 15095 bp in length
*	46326	46425: gap of unknown length
*	46426	50179: contig of 3754 bp in length
*	50180	50179: gap of unknown length
*	50280	91627: contig of 41348 bp in length
*	91628	91727: gap of unknown length
*	91728	112791: contig of 21064 bp in length
*	112792	112891: gap of unknown length
*	112892	124370: contig of 11479 bp in length
*	124371	124470: gap of unknown length
*	124471	128791: contig of 4321 bp in length
*	128792	128891: gap of unknown length
*	128892	153868: contig of 24977 bp in length
*	153869	153968: gap of unknown length
*	153969	195116: contig of 41148 bp in length
*	195117	195216: gap of unknown length
*	195217	221812: contig of 26596 bp in length
*	221813	221912: gap of unknown length
*	221913	228780: contig of 6868 bp in length
*	228781	228880: gap of unknown length
*	228881	255128: contig of 26248 bp in length.

FEATURES
 Location/Qualifiers
 1..255128
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="15"
 /map="15"
 /clone="RP24-303N12"
 /clone_lib="RPC1-24 Male Mouse BAC"

ORIGIN
 Query Match 68.0%; Score 20.4; DB 2; Length 255128;
 Best Local Similarity 80.0%; Pred No. 3.3e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps

QY	1	CTGGATCTTCGCCCGCCCGCCCTTGAATTC	30
Db	129187	CTGGTTCTTCCTTCCTGCCCGCCAGAAATCCC	129216

RESULT 43
AC098660/c

LOCUS	AC098660	264178 bp	DNA	linear	HTG 10-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-103B11, WORKING DRAFT SEQUENCE.				
ACCESSION	AC098660				
VERSION	6 GI:30520686				
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
REFERENCE	1 (bases 1 to 264178)				
AUTHORS	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, H., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X. -Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 264178)				
AUTHORS	Worley, K. C.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 264178)				
AUTHORS	Rat Genome Sequencing Consortium.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One				
COMMENT	<p>Baylor Plaza, Houston, TX 77030, USA</p> <p>On May 10, 2003 this sequence version replaced gi:25091776. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.</p> <p>----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GHWD Center clone name: CH230-103B11 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 236157 bases at least Q40 Consensus quality: 238656 bases at least Q30 Consensus quality: 240591 bases at least Q20 Estimated insert size: 249685; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation</p> <p>----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submitter. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. * * 1 264178: contig of 264178 bp in length.</p> <p>FEATURES source 1 .264178 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-103B11" 129216..131335 /note="wgs_contig"</p> <p>misc_feature 1 CTGGATCTTGGCCGCCCTTGAAATCCC 30 161402 CTGGATCTTAGCCCGCCCTTGATTTCC 161373 </p> <p>ORIGIN Query Match 68.0%; Score 20.4; DB 2; Length 264178; Best Local Similarity 80.0%; Pred. No. 3.3e+02; Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;</p> <p>QY 1 CTGGATCTTGGCCGCCCTTGAAATCCC 30 DB 161402 CTGGATCTTAGCCCGCCCTTGATTTCC 161373 </p> <p>RESULT 44 AC119085 LOCUS DEFINITION AC119085 AC119085.3 GI:23915363 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED. KEYWORDS Rattus norvegicus SOURCE Rattus norvegicus ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;</p>				

REFERENCE

1 (bases 1 to 264543)
 Muzny, D. Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceaar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Paotelmen, O., Okwuonu, G., Olampunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trefos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 264543)

Worley, K. C.

Direct Submission

Submitted (25-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 264543)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 12, 2002 this sequence version replaced gi:21745858.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center Project name: GINQ
 Center clone name: CH230-223G14

----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 237681 bases at least Q40
 Consensus quality: 239908 bases at least Q30
 Consensus quality: 241426 bases at least Q20
 Estimated insert size: 245780; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1	6181:	contig of 6181 bp in length
* 6182	6281:	gap of unknown length
* 6282	8969:	contig of 2688 bp in length
* 8970	9069:	gap of unknown length
* 9070	13423:	contig of 4354 bp in length
* 13424	13523:	gap of unknown length
* 13524	209535:	contig of 196012 bp in length
* 209536	209635:	gap of unknown length
* 209636	21318:	contig of 5883 bp in length
* 21319	213418:	gap of unknown length
* 215419	24340:	contig of 32922 bp in length
* 248341	248440:	gap of unknown length
* 248441	250634:	contig of 2194 bp in length
* 250635	250734:	gap of unknown length
* 250735	264543:	contig of 13809 bp in length.

FEATURES

source

1..264543
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-223G14"

1..1047

/note="wgs_end_extension
 clone_end:T7"

321..1232

/note="clone_boundary

clone_end:T7

site:ECORI

end_sequence:RWBNB43TJB"

2568..4314

/note="wgs_contig"

4365..6181

/note="wgs_contig"

6282..7426

/note="wgs_contig"

7606..8969

/note="wgs_contig"

9070..10259

/note="wgs_contig"

10741..13423

/note="wgs_contig"

209636..211486

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
/note="wgs_contig"
misc_feature 211537..213081
/note="wgs_contig"
misc_feature 213132..215318
/note="wgs_contig"

ORIGIN
Query Match 68.0%; Score 20.4; DB 2; Length 264543;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGGATCCTTGGCCCGCCCTTGAATTCCC 30
|||||
Db 191881 CTGCATCATTTGCCACCTTCAATCCC 191910
|||||

RESULT 45
AE016861/c
LOCUS
DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 6 of 21 of the
complete genome.
ACCESSION AE016861
VERSION AE016861.1 GI:28851889
KEYWORDS
SOURCE Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Pseudomonas syringae pv. tomato str. DC3000
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 310029)
AUTHORS Buell,C.R., Joardar,V., Lindeberg,M., Selengut,J., Paulsen,I.T.,
Gwinn,M.L., Dodson,R.J., DeBoy,R.T., Durkin,A.S., Kolonay,J.F.,
Madupu,R., Daugherty,S., Brinkac,L., Beanan,M.J., Haft,D.H.,
Nelson,W.C., DavidSEN,T., Zafar,N., Zhou,L., Liu,J., Yuan,Q.,
Khouri,H., Fedorova,N., Tran,B., Russell,D., Berry,K.,
Utterback,T., Van Aken,S.E., Feldblyum,T.V., D'Ascenzo,M.,
Deng,W.L., Rancos,A.R., Alfano,J.R., Cartinhour,S., Chatterjee,A.K.,
Delaney,T.P., Lazarowitz,S.G., Martin,G.B., Schneider,D.J.,
Tang,X., Bender,C.L., White,O., Fraser,C.M. and Collmer,A.
The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000
Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
12928499

JOURNAL
PUBMED
REFERENCE 2 (bases 1 to 310029)
AUTHORS Buell,R., Joardar,V., Khouri,H., Fedorova,N., Tran,B., Russell,D.,
Berry,K., Utterback,T., Van Aken,S., Feldblyum,T., Gwinn,M.,
Dodson,R., DeBoy,R., Durkin,A., Kolonay,J., Madupu,R.,
Daugherty,S., Brinkac,L., Beanan,M., Haft,D., Selengut,J.,
Nelson,W., DavidSEN,T., White,O., Fraser,C. and Collmer,A.
Direct Submission
Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
Location/Qualifiers
source 1..310029
/organism="Pseudomonas syringae pv. tomato str. DC3000"
/mol_type="genomic DNA"
/strain="DC3000"
/db_xref="taxon:232283"
/note="pathovar: tomato"
complement(217..1974)
/gene="leuA"
/locus_tag="PSPT01444"
/codon_start=1
/transl_table=1
/product="2-isopropylmalate synthase"
/protein_id="AAO54965.1"
/db_xref="GI:28851890"
/translation="MPPRCVRYPHVFTAGRLVPAPSLQKDFQMTMLNDPSKKYRAFP
TIDLPRTPSKTIDAPWCSLDGNSLIEPWSVKLRFLFKTLVSVGKEIEA
SPPAASQTFDFRTLIEDGHIPDDTTIQVLTQAREDLIARTFESLRGAKKAIVHLN
ATCPSFRRIYFNQDKAGVKEIAVNAKLFVKYAAQQPQTQWTFEYSPETFSLELEFA

gene
/locus_tag="PSPT01444"
complement(217..1974)
/gene="leuA"
/locus_tag="PSPT01444"
/codon_start=1
/transl_table=1
/product="2-isopropylmalate synthase"
/protein_id="AAO54965.1"
/db_xref="GI:28851890"
/translation="MPPRCVRYPHVFTAGRLVPAPSLQKDFQMTMLNDPSKKYRAFP
TIDLPRTPSKTIDAPWCSLDGNSLIEPWSVKLRFLFKTLVSVGKEIEA
SPPAASQTFDFRTLIEDGHIPDDTTIQVLTQAREDLIARTFESLRGAKKAIVHLN
ATCPSFRRIYFNQDKAGVKEIAVNAKLFVKYAAQQPQTQWTFEYSPETFSLELEFA

CDS
5412..6881
/gene="guab"
/locus_tag="PSPT01449"
5412..6881
/gene="guab"
/locus_tag="PSPT01449"
/codon_start=1
/transl_table=11

KEYCDAVIEVWNPTEKNKVLNLPATVEVATNIVADQIEMEGRHITREDSVLISLHT
HNDRTGVAATELGMAGADRVEGCLFNGERTGNVDLVITVALNLTQGINBELRSD
IDGVRKVEECNQIFVPRHPYVGLDVHTAFSGSHQDAIRKGTQKQEGELMEVPYLP
IDPADIGRSYEAIVRNSQSGKGIAYLLEQYIGISLPRMQIBFSQVQVGETRLGL
EMTAQIHALLRREYLOANTPYALIASHRLOEENGNSVDAEVHVDGETQHRGKGKA
LEALVAGLPVAVEIMDYNEHAIGSGTTAKAAAYIELRVNGDRAVHGVDIDENTITASF
RALFSALNRSLSQAQKAA"
complement(2268..3092)
/locus_tag="PSPT01445"
complement(2268..3092)
/locus_tag="PSPT01445"
/codon_start=1
/transl_table=11
/product="peptidase, M23/M37 family"
/protein_id="AAO54966.1"
/db_xref="GI:28851891"
/translation="MLRTIAPLFAILLCLPAHADSPITRALDKPVGVAVIDLGTGA
CAPATYQKPVLVVKEQGTWLAIVGIPLVKPGTQOVTSGRTMSFTVGSKYPEQ
HITLKNQVNPENEDNKRIEGELAEQIRAVSFSPGTPSNLILDKPVNGPLSSKFGV
RRFNGERNPHSGLDPAVPACTPIKSPAGKVIITGNYPFNGNTVFPVDHGQGFISMF
CHLSKIDVKGQVPRGVGVKVGATGRATGPHMHNVSNDARVDPAIFIGAFQP"
complement(3099..4490)
/gene="xseA"
/locus_tag="PSPT01446"
complement(3099..4490)
/gene="xseA"
/locus_tag="PSPT01446"
/codon_start=1
/transl_table=11
/product="exodeoxyribonuclease VII, large subunit"
/protein_id="AAO54967.1"
/db_xref="GI:28851892"
/translation="WMLAMIKDPFARLGLDRELVTSQNGRARVLLEDVFSIWEVG
EISNLSRPSAGHYFTLKDSQAVRCALFRQSAARVQALKDGLQVKGKVSLEFGR
GDYQILITDTPAGDGLALFADLAKLSDEGLFSADRKVALPLHPQRIIGIISSTG
AVIRIDIIVERRARPVELTIPTAQCGREALNQIVRALKLADAGFDALILARGGGS
LEDLWCFNEEAVARALDCACTPVSVAAGHETDVSISDFVADVPATPSAAELLAPDS
SDLHRRVNHLHRLVSRMODRLMRERLRLEISRLRHPGERLRQOSQRDLDDMLRL
RAFESMKHKKQVRLAHMESRTAQHPGRTLAFRLQRLDALAERLPRATREQIKVKLKQ
LQSQMTLVVSPLATLGRGYSILLDERGHAIRNAOQTPQORLTARLEGELQVRVE
DNHLTPVTLSLLD"
4505..4693
/locus_tag="PSPT01447"
4505..4693
/locus_tag="PSPT01447"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAO54968.1"
/db_xref="GI:28851893"
/translation="MGASLLLTGTSTTSLAALRPPLLGLSLAGRALIKKLRSFAFVRLV
TWLVVSGLALIVRYLTA"
4769..5317
/locus_tag="PSPT01448"
4769..5317
/locus_tag="PSPT01448"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO54969.1"
/db_xref="GI:28851894"
/translation="MNTOSIIVPQISTPFGHEARARLILRWLVKLDVPELTTCORT
YNKMYAVAPGARVRYKKEALPGQVINGLEIVTKRCIFTPLNPAEAGGPECCHRE
VGEALFDSLEDWMPGHTDNFTCECRHEDDINGFLDCAAGFSLNIGFIINNLDASFT
QTFLDDFAERLDRPVSVFHVRL"
5412..6881
/gene="guab"
/locus_tag="PSPT01449"
5412..6881
/gene="guab"
/locus_tag="PSPT01449"
/codon_start=1
/transl_table=11
```

/product="inosine-5-monophosphate dehydrogenase"
/protein_id="AA054970.1"
/db_xref="GI:28851895"
/translation="MLRIKQKALTFDDILLVPGYSEVLPLEVSLKTRLTRGIELNIPL
VSAAMDTVEARLATAAGGEGIGIIHKMTIEQQAEEVRKVKKEAGVVKDPITIEA
DATVRLDFLEPHNIGSVFVLHNGDLGVITSDRVFENRLDVPVEMTPKERLVT
VREGADKNEVRELLHRLKLEKLIVDANFALGMMTKVDEIKAKAYPLASQDQARLR
VGAAGTGDCTGERTALVAAGVDVVVDTAHGHSGVTDVRVWVKDNFPQVQVIGGN
IATGEAALVAAGADAVKVGIGPGSICTTRIVAGVGVPQISAIANVAALAEGLGVP
IADGIRSGDLSKALVAGASCVMGSMFAGTEAPGEIELEFQGSYKAYRGMGSLGA
MSQAQSGSDRYQDSAGAKLVPGIEGRVAIKGPLSAIHQLMGLGRSSNGYTGSA
DIEQMRITPEFVRIITGAGNASHVHDVQITKEAPNRYG"
7005. .8582
/gene="guaA"
/locus_tag="PSPTO1450"
7005. .8582
/gene="guaA"
/locus_tag="PSPTO1450"
/codon_start=1
/transl_table=11
/product="GMP synthase"
/protein_id="AA054971.1"
/db_xref="GI:28851896"
/translation="WALDIHAHRIILDFGSQYTLIARRVREIGVYCELPFDMDE
AIREPAPKGVILAGPESVHEANSRCPQAVFDLGVFVFGICYGQMTWAEOLGKVGAG
SELREFGAVDVGKSLRDGIEDHIDADGLFDLVMHSGDKVKTLKLPEDPHILAST
PSRCFAGWADRVGQYGFPEVHTTKGGILSRFLIDICGCEALWTPSKIAEDAI
AQVRAQGTDSVLGSGDSSVVAALLHKAIGDLQICVFDVNGLLRLHGEQVWAM
FAENMGVIRANAEQKFLNNLESGDPEKKRIIGRTFIDVDFDAESCKJONIKYLAQ
GTIYDPDIESAGAKGVKSHNVGGLPEENMLKLVPELRFELDFEVRRLGLELG
LPDVMVRRHPGPGVGLGVEKVEYADILRADHIFIEELRKADWYHKVQAFVV
FOPKSVGVGGRVYAVVALRAVETIDFMTARWAHLPYELLETVSGRINEIGIS
RVTYDVSSKPPATIEWE"
complement(9533. .10714)
/locus_tag="PSPTO1451"
complement(9533. .10714)
/locus_tag="PSPTO1451"
/note="similar to GP:17982579; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA054972.1"
/db_xref="GI:28851897"
/translation="WATANKLKATWATVTLNLDWPSASLLTIDEARVSTOKEELT
GLIDIVAMKMWAVIQVPCADALYASDLLPWSKYLTGTLGNKFGFDPFLAYVQQA
ARHIELHAWNVYRISMNASDGTMBELNNSSDSPASVFNTHPEWTGAANRFLVLPNG
IPEVQAWGSIYVEETVYVDVAIQDDYFYFETADSLQDDATYTYNTFTTAKADW
RNNYTSILDACHKXIAAVTDVLFGISGAPVWRNKSDDPLGSDTQAGAPNYDFAYAD
TRKWIDGIDYIAPOVYVPPFAREAVRDVLTQWADTVRGTGTYLYGVGMALYKVGAA
SNAEPDWTVEAGVPENTQLDNLDSLAESVGCMLFRHMFELRASQTQVVYDVKLRWAD
V"
complement(10803. .10919)
/locus_tag="PSPTO1452"
complement(10803. .10919)
/locus_tag="PSPTO1452"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA054973.1"
/db_xref="GI:28851898"
/translation="MMPKRLCEQALFVRSQAQLAHRDNLILGYGHIPAPWQV"
complement(10934. .12229)
/gene="lec-1"
/locus_tag="PSPTO1453"
complement(10934. .12229)
/gene="lec-1"
/locus_tag="PSPTO1453"
/note="similar to SP:O68609, and SP:O68609; identified by
sequence similarity; putative"
/codon_start=1

Query Match 68.0%; Score 20.4; DB 1; Length 310029;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGATCCTTGGCCCGCCCTTGAATTC 30
DB 273065 CTGAATCCTTGGCCCTGCTTTAATTC 273036
RESULT 46
AP000060/c
LOCUS Aeropyrum pernix K1 DNA, complete genome, section 3/7.
DEFINITION AP000060 BA000002
ACCESSION AP000060.1 GI:5104188
VERSION
KEYWORDS
SOURCE Aeropyrum pernix K1
ORGANISM Aeropyrum pernix K1
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaceae; Aeropyrum.
REFERENCE 1
AUTHORS Kwarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y.,
Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Anka, A., Kosugi, H.,
Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H.,
Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y.,
Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K.,
Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H.
Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1
DNA Res. 6 (2), 83-101 (1999)
99310339
PUBMED 10382966
REFERENCE 2 (bases 1 to 347800)
AUTHORS Director-General of Biotechnology Center, Tanaka, T., Hino, Y.,
Kwarabayasi, Y. and Kikuchi, H.
Direct Submission
Submitted (14-DEC-1998) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center, 2-49-10 Nishihara, Shibuyaku, Tokyo 151-0066, Japan
(E-mail:bio@nitech.go.jp, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
Kwarabayasi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
606-8502, Japan
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 109; EMBL release 56.0; SwissProt
release 36.0; PIR-Protein release 57.0; and OWL release 31.0.
E-mail address for comments and questions: genome@nitech.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: http://www.nitech.go.jp/.
FEATURES
source
1. .347800
/organism="Aeropyrum pernix K1"
/mol_type="genomic DNA"
/strain="K1"
/db_xref="taxon:272557"
complement(145. .232)
/gene="APECRNA12"
complement(145. .232)
/gene="APECRNA12"
/product="crRNA-Leu"
/note="codon recognized: CUA; anticodon:TAG"
/anticodon=(pos:194. .196,aa:Leu)
600. .1574
/gene="APE0537"
600. .1574
/gene="APE0537"
CDS

RESULT 47

AC100623/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

AC100623 58554 bp DNA linear HTG 22-NOV-2001
Mus musculus clone RP23-161B20, LOW-PASS SEQUENCE SAMPLING.

AC100623

AC100623.1 GI:17047989

HTG; HTGS PHASE0.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 58554)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-161B20

Unpublished

2 (bases 1 to 58554)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L15814

Center clone name: 161_B_20

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 681: contig of 681 bp in length
* 682 781: gap of 100 bp
* 782 1494: contig of 713 bp in length
* 1495 1594: gap of 100 bp
* 1595 2320: contig of 726 bp in length
* 2321 2420: gap of 100 bp
* 2421 3120: contig of 700 bp in length

* 3121 3220: gap of 100 bp
* 3221 3921: contig of 701 bp in length
* 3922 4021: gap of 100 bp
* 4022 4705: contig of 684 bp in length
* 4706 4805: gap of 100 bp
* 4806 5454: contig of 649 bp in length
* 5455 5554: gap of 100 bp
* 5555 6222: contig of 668 bp in length
* 6223 6322: gap of 100 bp
* 6323 6993: contig of 671 bp in length
* 6994 7093: gap of 100 bp
* 7094 7792: contig of 699 bp in length
* 7793 7892: gap of 100 bp
* 7893 8610: contig of 718 bp in length
* 8611 9427: contig of 717 bp in length
* 9428 9527: gap of 100 bp
* 9528 10226: contig of 699 bp in length
* 10227 10326: gap of 100 bp
* 10327 11039: contig of 713 bp in length
* 11040 11139: gap of 100 bp
* 11140 11824: contig of 685 bp in length
* 11825 11924: gap of 100 bp
* 11925 12621: contig of 697 bp in length
* 12622 12721: gap of 100 bp
* 12722 13453: contig of 732 bp in length
* 13454 13553: gap of 100 bp
* 13554 14278: contig of 725 bp in length
* 14279 14378: gap of 100 bp
* 14379 15063: contig of 685 bp in length
* 15064 15163: gap of 100 bp
* 15164 15908: contig of 745 bp in length
* 15909 16008: gap of 100 bp
* 16009 16742: contig of 734 bp in length
* 16743 16842: gap of 100 bp
* 16843 17549: contig of 707 bp in length
* 17550 17649: gap of 100 bp
* 17650 18305: contig of 656 bp in length
* 18306 18405: gap of 100 bp
* 18406 19104: contig of 699 bp in length
* 19105 19204: gap of 100 bp
* 19205 19903: contig of 699 bp in length
* 19904 20003: gap of 100 bp
* 20004 20672: contig of 669 bp in length
* 20673 20772: gap of 100 bp
* 20773 21452: contig of 680 bp in length
* 21453 21552: gap of 100 bp
* 21553 22280: contig of 728 bp in length
* 22281 22380: gap of 100 bp
* 22381 23123: contig of 743 bp in length
* 23124 23223: gap of 100 bp
* 23224 23914: contig of 691 bp in length
* 23915 24014: gap of 100 bp
* 24015 24751: contig of 737 bp in length
* 24752 24851: gap of 100 bp
* 24852 25558: contig of 707 bp in length
* 25559 25658: gap of 100 bp
* 25659 26365: contig of 707 bp in length
* 26366 26465: gap of 100 bp
* 26466 27158: contig of 693 bp in length
* 27159 27258: gap of 100 bp
* 27259 27938: contig of 680 bp in length
* 27939 28038: gap of 100 bp
* 28039 28709: contig of 671 bp in length
* 28710 28809: gap of 100 bp
* 28810 29499: contig of 690 bp in length
* 29500 29599: gap of 100 bp
* 29600 30292: contig of 693 bp in length
* 30293 30392: gap of 100 bp
* 30393 31138: contig of 746 bp in length
* 31139 31238: gap of 100 bp
* 31239 31939: contig of 701 bp in length
* 31940 32039: gap of 100 bp

* 32040 32722: contig of 683 bp in length
 * 32723 32822: gap of 100 bp
 * 32823 33566: contig of 744 bp in length
 * 33567 33666: gap of 100 bp
 * 33667 34397: contig of 731 bp in length
 * 34398 34497: gap of 100 bp
 * 34498 35199: contig of 702 bp in length
 * 35200 35299: gap of 100 bp
 * 35300 35988: contig of 689 bp in length
 * 35989 36088: gap of 100 bp
 * 36089 36777: contig of 689 bp in length
 * 36778 36877: gap of 100 bp
 * 36878 37591: contig of 714 bp in length
 * 37592 37691: gap of 100 bp
 * 37692 38411: contig of 720 bp in length
 * 38412 38511: gap of 100 bp
 * 38512 39255: contig of 744 bp in length
 * 39256 39355: gap of 100 bp
 * 39356 40050: contig of 695 bp in length
 * 40051 40150: gap of 100 bp
 * 40151 40848: contig of 698 bp in length
 * 40849 40948: gap of 100 bp
 * 40949 41628: contig of 680 bp in length
 * 41629 41728: gap of 100 bp
 * 41729 42413: contig of 685 bp in length
 * 42414 42513: gap of 100 bp
 * 42514 43223: contig of 710 bp in length
 * 43224 43323: gap of 100 bp
 * 43324 44007: contig of 684 bp in length
 * 44008 44107: gap of 100 bp
 * 44109 44935: contig of 728 bp in length
 * 44936 44935: gap of 100 bp
 * 44936 45660: contig of 725 bp in length
 * 45661 45760: gap of 100 bp
 * 45761 46464: contig of 704 bp in length
 * 46465 46564: gap of 100 bp
 * 46565 47297: contig of 733 bp in length
 * 47298 47397: gap of 100 bp
 * 47398 48116: contig of 719 bp in length
 * 48117 48217: gap of 100 bp
 * 48217 48917: contig of 701 bp in length
 * 48918 49017: gap of 100 bp
 * 49018 49728: contig of 711 bp in length
 * 49729 49828: gap of 100 bp
 * 49829 50497: contig of 669 bp in length
 * 50498 50597: gap of 100 bp
 * 50598 51305: contig of 708 bp in length
 * 51306 51405: gap of 100 bp
 * 51406 52083: contig of 678 bp in length
 * 52084 52183: gap of 100 bp
 * 52184 52862: contig of 679 bp in length
 * 52863 52962: gap of 100 bp
 * 52963 53708: contig of 745 bp in length
 * 53709 53808: gap of 100 bp
 * 53809 54598: contig of 690 bp in length
 * 54599 54598: gap of 100 bp
 * 54599 55315: contig of 717 bp in length
 * 55316 55415: gap of 100 bp

Query Match 67.3%; Score 20.2; DB 2; Length 58554;
 Best Local Similarity 88.0%; Pred. No. 4.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 6 TCCTTGCCTCCGCTTGAATTTCCC 30
 Db 47297 TCCTTGCCTCCGCTTGAATTTCCC 47273

RESULT 48

AC112449 201376 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-241H9, *** SEQUENCING IN PROGRESS
 DEFINITION *** 2 unordered pieces.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC112449 GI:25072616
 HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS

1 (bases 1 to 201376)
 Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barbee, M., Barnstead, M., Benahmed, F.,
 Bryant, N., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuwa, L., Loulsee, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajda, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Ugmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstein, G. and Gibbs, R. A.

TITLE
JOURNAL

Unpublished
 2 (bases 1 to 201376)
 Worley, K. C.
 Direct Submission

TITLE
 JOURNAL

Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

3 (bases 1 to 201376)
 Rat Genome Sequencing Consortium.
 Direct Submission

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23817380.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRYZ
Center clone name: CH230-241H9

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 161070 bases at least Q40
Consensus quality: 162780 bases at least Q30
Consensus quality: 164072 bases at least Q20
Estimated insert size: 168831; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 198501: contig of 198501 bp in length
* 198502 198601: gap of unknown length
* 198602 201376: contig of 2775 bp in length.

FEATURES

source

1. 201376
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-241H9"

misc_feature

1. 1749
/note="wgs_end_extension"

misc_feature

4975..6734
/note="wgs_end_extension"

misc_feature

clone_end:T7"
complement(8569..9431)
/note="clone_boundary"

misc_feature

clone_end:T7
site:
end_sequence:BT151820"
197495..198501
/note="wgs_contig"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 201376;
Best Local Similarity 88.0%; Pred. No. 4.1e-02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 6 TCCTTCCCGCCGCTTGAATCCC 30

DB 168181 TCCTTCCCGCTTGAATCCC 168205

RESULT 49

AC132971

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC132971
Rattus norvegicus clone CH230-125D24, WORKING DRAFT SEQUENCE.
AC132971
AC132971.4 GI:25007428
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 244693)
Murny,D., Marie., Metzker,M., Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswato,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C., Li,Z., Liu,J.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwalekeme,O., Okwuon,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umami,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 244693)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 244693)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT	Baylor Plaza, Houston, TX 77030, USA									
	On Nov 15, 2002 this sequence version replaced gi:23196084. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.									
	----- Genome Center									
	Center: Baylor College of Medicine									
	Center code: BCM									
	Web site: http://www.hgsc.bcm.tmc.edu/									
	Contact: hgsc-help@bcm.tmc.edu									
	----- Project Information									
	Center project name: KBLZ									
	Center clone name: CH230-125D24									
----- Summary Statistics										
Assembly program: Phrap; version 0.990329										
Consensus quality: 219156 bases at least Q40										
Consensus quality: 222417 bases at least Q30										
Consensus quality: 224165 bases at least Q20										
Estimated insert size: 224807; sum-of-contigs estimation										
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation										

* NOTE: Estimated insert size may differ from sequence length										
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).										
* NOTE: This is a 'working draft' sequence. It currently										
* consists of 1 contigs. Gaps between the contigs										
* are represented as runs of N. The order of the pieces										
* is believed to be correct as given, however the sizes										
* of the gaps between them are based on estimates that have										
* provided by the submitter.										
* This sequence will be replaced										
* by the finished sequence as soon as it is available and										
* the accession number will be preserved.										
* 1 244693: contig of 244693 bp in length.										
Location/Qualifiers										
1. .244693										
/organism="Rattus norvegicus"										
/mol_type="genomic DNA"										
/db_xref="taxon:10116"										
/clone="CH230-125D24"										
1. .1013										
/notes="wgs_end_extension										
clone_end:T7"										
1186-.12882										
/notes="wgs_end_extension										
clone_end:T7"										
17826-.18076										
/notes="clone_boundary										
clone_end:T7"										
site:										
end_sequence:RWAUT24TJC"										
ORIGIN										
Query Match 67.3%; Score 20.2; DB 2; Length 244693;										
Best Local Similarity 88.0%; Pred. No. 4e+02; 3; Indels 0;										
Matches 22; Conservative 0; Mismatches 0; Gaps 0;										
Qy 6 TCCTTCCCGCCGCTTGAATCCC 30										
Db 122971 TCCTTCCCGCCGCTTGAATCCC 122995										
AC120747 255646 bp DNA linear HTG 09-OCT-2002										
RESULT 50										
AC120747										
LOCUS										

COMMENT	Baylor Plaza, Houston, TX 77030, USA									
	On Nov 15, 2002 this sequence version replaced gi:23196084.									
	The sequence in this assembly is a combination of BAC based reads									
	and whole genome shotgun sequencing reads assembled using Atlas									
	(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described									
	in the feature table below represents a scaffold in the Atlas									
	assembly (a 'contig-scaffold'). Within each contig-scaffold,									
	individual sequence contigs are ordered and oriented, and separated									
	by sized gaps filled with Ns to the estimated size. The sequence									
	may extend beyond the ends of the clone and there may be sequence									
FEATURES	contigs within a contig-scaffold that consist entirely of whole									
	genome shotgun sequence reads. Both end sequences and whole genome									
	shotgun sequence only contigs will be indicated in the feature									
	table.									
	----- Genome Center									
	Center: Baylor College of Medicine									
	Center code: BCM									
	Web site: http://www.hgsc.bcm.tmc.edu/									
	Contact: hgsc-help@bcm.tmc.edu									
	----- Project Information									
Center project name: KBLZ										
Center clone name: CH230-125D24										
----- Summary Statistics										
Assembly program: Phrap; version 0.990329										
Consensus quality: 219156 bases at least Q40										
Consensus quality: 222417 bases at least Q30										
Consensus quality: 224165 bases at least Q20										
Estimated insert size: 224807; sum-of-contigs estimation										
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation										

* NOTE: Estimated insert size may differ from sequence length										
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).										
* NOTE: This is a 'working draft' sequence. It currently										
* consists of 1 contigs. Gaps between the contigs										
* are represented as runs of N. The order of the pieces										
* is believed to be correct as given, however the sizes										
* of the gaps between them are based on estimates that have										
* provided by the submitter.										
* This sequence will be replaced										
* by the finished sequence as soon as it is available and										
* the accession number will be preserved.										
* 1 244693: contig of 244693 bp in length.										
Location/Qualifiers										
1. .244693										
/organism="Rattus norvegicus"										
/mol_type="genomic DNA"										
/db_xref="taxon:10116"										
/clone="CH230-125D24"										
1. .1013										
/note="wgs_end_extension"										
clone_end:T7"										
11186..12882										
/note="wgs_end_extension"										
clone_end:T7"										
17826..18076										
/note="clone boundary"										
clone_end:T7"										
site:										
end_sequence:RWAUT24TJC"										
ORIGIN										
Query Match 67.3%; Score 20.2; DB 2; Length 244693;										
Best Local Similarity 88.0%; Pred. No. 4e+02;										
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;										
Qy 6 TCCTTCCCGCCGCTTGAATCCC 30										
Db 122971 TCCTTCCCGCTTGAATCCC 122995										
RESULT 50										
AC120747										
LOCUS										

DEFINITION	Rattus norvegicus clone CH230-25J5, *** SEQUENCING IN PROGRESS ***
ACCESSION	AC120747
VERSION	AC120747.4 GI:23269576
KEYWORDS	HTG: HTGS; PASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	1 (bases 1 to 255646)
AUTHORS	Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshuwa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwackelmech,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.
REFERENCE	2 (bases 1 to 255646)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 255646)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	

Baylor Plaza, Houston, TX 77030, USA	On Nov 15, 2002 this sequence version replaced gi:23196084. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
Center: Baylor College of Medicine	Genome Center
Center code: BCM	Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/	Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu	Contact: hgsc-help@bcm.tmc.edu
----- Project Information	----- Project Information
Center project name: KBIZ	Center project name: KBIZ
Center clone name: CH230-125D24	Center clone name: CH230-125D24
----- Summary Statistics	----- Summary Statistics
Assembly program: Phrap; version 0.990329	Assembly program: Phrap; version 0.990329
Consensus quality: 219156 bases at least Q40	Consensus quality: 219156 bases at least Q40
Consensus quality: 22417 bases at least Q30	Consensus quality: 22417 bases at least Q30
Consensus quality: 224165 bases at least Q20	Consensus quality: 224165 bases at least Q20
Estimated insert size: 224807; sum-of-contigs estimation	Estimated insert size: 224807; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation	Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----	-----
* NOTE: Estimated insert size may differ from sequence length	* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).	(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently	* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs	consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces	are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes	is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have	of the gaps between them are based on estimates that have
provided by the submitter.	provided by the submitter.
* This sequence will be replaced	* This sequence will be replaced
by the finished sequence as soon as it is available and	by the finished sequence as soon as it is available and
the accession number will be preserved.	the accession number will be preserved.
* 1 244693: contig of 244693 bp in length.	* 1 244693: contig of 244693 bp in length.
Location/Qualifiers	Location/Qualifiers
1. .244693	1. .244693
/organism="Rattus norvegicus"	/organism="Rattus norvegicus"
/mol_type="genomic DNA"	/mol_type="genomic DNA"
/db_xref="taxon:10116"	/db_xref="taxon:10116"
/clone="CH230-125D24"	/clone="CH230-125D24"
1. .1013	1. .1013
/note="wgs_end_extension"	/note="wgs_end_extension"
clone_end:T7"	clone_end:T7"
11186..12882	11186..12882
/note="wgs_end_extension"	/note="wgs_end_extension"
clone_end:T7"	clone_end:T7"
17826..18076	17826..18076
/note="clone boundary"	/note="clone boundary"
clone_end:T7"	clone_end:T7"
site:	site:
end_sequence:RWAUT24TJC"	end_sequence:RWAUT24TJC"
ORIGIN	ORIGIN
Query Match 67.3%; Score 20.2; DB 2; Length 244693;	Query Match 67.3%; Score 20.2; DB 2; Length 244693;
Best Local Similarity 88.0%; Pred. No. 4e+02; 0; Gaps 0;	Best Local Similarity 88.0%; Pred. No. 4e+02; 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0;	Matches 22; Conservative 0; Mismatches 3; Indels 0;
Qy 6 TCCTTCCCGCCCGCTGAATCCC 30	Qy 6 TCCTTCCCGCCCGCTGAATCCC 30
Db 122971 TCCTTCCCGCTCCCTTAATCCC 122995	Db 122971 TCCTTCCCGCTCCCTTAATCCC 122995
RESULT 50	RESULT 50
AC120747	AC120747
LOCUS	LOCUS

Baylor Plaza, Houston, TX 77030, USA	On Nov 15, 2002 this sequence version replaced gi:23196084. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
Center: Baylor College of Medicine	Center: Baylor College of Medicine
Center code: BCM	Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/	Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu	Contact: hgsc-help@bcm.tmc.edu
----- Project Information	----- Project Information
Center project name: KBLZ	Center project name: KBLZ
Center clone name: CH230-125D24	Center clone name: CH230-125D24
----- Summary Statistics	----- Summary Statistics
Assembly program: Phrap; version 0.990329	Assembly program: Phrap; version 0.990329
Consensus quality: 219156 bases at least Q40	Consensus quality: 219156 bases at least Q40
Consensus quality: 222417 bases at least Q30	Consensus quality: 222417 bases at least Q30
Consensus quality: 224165 bases at least Q20	Consensus quality: 224165 bases at least Q20
Estimated insert size: 224807; sum-of-contigs estimation	Estimated insert size: 224807; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation	Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----	-----
* NOTE: Estimated insert size may differ from sequence length	* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).	(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently	* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs	consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces	are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes	is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have	of the gaps between them are based on estimates that have
provided by the submitter.	provided by the submitter.
* This sequence will be replaced	* This sequence will be replaced
* by the finished sequence as soon as it is available and	* by the finished sequence as soon as it is available and
* the accession number will be preserved.	* the accession number will be preserved.
* 1 244693: contig of 244693 bp in length.	* 1 244693: contig of 244693 bp in length.
Location/Qualifiers	Location/Qualifiers
1. .244693	1. .244693
/organism="Rattus norvegicus"	/organism="Rattus norvegicus"
/mol_type="genomic DNA"	/mol_type="genomic DNA"
/db_xref="taxon:10116"	/db_xref="taxon:10116"
/clone="CH230-125D24"	/clone="CH230-125D24"
1. .1013	1. .1013
/note="wgs_end_extension"	/note="wgs_end_extension"
clone_end:T7	clone_end:T7
1186..12882	1186..12882
/note="wgs_end_extension"	/note="wgs_end_extension"
clone_end:T7	clone_end:T7
17826..18076	17826..18076
/note="clone_boundary"	/note="clone_boundary"
clone_end:T7	clone_end:T7
site:	site:
end_sequence:RWAUT24TJC"	end_sequence:RWAUT24TJC"
ORIGIN	ORIGIN
Query Match 67.3%; Score 20.2; DB 2; Length 244693;	Query Match 67.3%; Score 20.2; DB 2; Length 244693;
Best Local Similarity 88.0%; Pred. NO. 4e+02;	Best Local Similarity 88.0%; Pred. NO. 4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 6 TCCTTCCCGCCGCTTGAATCCC 30	Qy 6 TCCTTCCCGCCGCTTGAATCCC 30
Db 122971 TCCTTCCCGCTTGAATCCC 122995	Db 122971 TCCTTCCCGCTTGAATCCC 122995
RESULT 50	RESULT 50
AC120747	AC120747
LOCUS	LOCUS

COMMENT

Baylor Plaza, Houston, TX 77030, USA
 On Sep 23, 2002 this sequence version replaced gi:21908095.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXYW
 Center clone name: CH230-25J5
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 214408 bases at least Q40
 Consensus quality: 218596 bases at least Q30
 Consensus quality: 220431 bases at least Q20
 Estimated insert size: 243823; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 73648: contig of 73648 bp in length
 * 73649 73748: gap of unknown length
 * 73749 84455: contig of 10707 bp in length
 * 84456 84555: gap of unknown length
 * 84556 253212: contig of 168657 bp in length
 * 253213 253312: gap of unknown length
 * 253313 254433: contig of 1121 bp in length
 * 254434 254533: gap of unknown length
 * 254534 255646: contig of 1113 bp in length.

FEATURES

source
 1..255646
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-25J5"
 misc_feature
 1..1581
 /note="wgs contig"
 misc_feature
 71900..73648
 /note="wgs contig"
 misc_feature
 73749..75116
 /note="wgs contig"
 misc_feature
 244993..246075
 /note="wgs contig"
 misc_feature
 251982..253212
 /note="wgs contig"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 255646;
 Best Local Similarity 88.0%; Pred. No. 4e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 6 TCCTTCCCCCGCCCTTGAATCCC 30
 |||||
 Db 100690 TCCTTCCCCCTCCCTTAATCCC 100714

Search completed: April 25, 2005, 12:17:22
 Job time : 680.103 secs